GenCore version 5.1.3

OM nucleic - nucleic search, using sw model

Run on: November 16, 2002, 23:19:15 ; Search time 4188 Seconds
(without alignments)
16601.386 Million cell updates/sec

Title: US-08-961-083-55
Perfect score: 2389
Sequence: 1 TTCTTACGAGTTGGGACTGT.....TAAGTAAGGAAAAAATAAAC 2389

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		2388	100.0	2389	6	AR120265	AR120265 Sequence
2		2388	100.0	2541	1	AF291695	AF291695 Streptoco
3		2388	100.0	8195	6	BD003774	BD003774 Polynucle
4	c	2388	100.0	10256	1	AE007418	AE007418 Streptoco
5	c	2365.6	99.0	10320	1	AE008479	AE008479 Streptoco
6		2328.2	97.5	20035	2	SPNEU1915	AL449937 Streptoco
7		1434.4	60.0	2535	1	AF340221	AF340221 Streptoco
8		1374.2	57.5	2523	6	AX343072	AX343072 Sequence
9		1374.2	57.5	2647	6	AX343073	AX343073 Sequence
10		991	41.5	2457	1	AF318954	AF318954 Streptoco
11		990	41.4	2517	1	AF318955	AF318955 Streptoco
12		990	41.4	11931	1	AE007403	AE007403 Streptoco
13		988.4	41.4	232807	2	SPNEU1901	AL449923 Streptoco
14		987.6	41.3	2290	6	AR120270	AR120270 Sequence
15		980.8	41.1	2639	6	AX343074	AX343074 Sequence
16		962.8	40.3	12372	1	AE008464	AE008464 Streptoco
17		869.4	36.4	2166	12	AF340222	AF340222 Synthetic
18	c	835.8	35.0	75874	2	SPNEU1907	AL449929 Streptoco
19		674	28.2	702	2	SPNEU1929	AL449951 Streptoco
20		653.8	27.4	2359	6	BD003923	BD003923 Polynucle
21		385.4	16.1	1342	6	AR120328	AR120328 Sequence
22		385.4	16.1	3120	1	AF318956	AF318956 Streptoco
23		385.4	16.1	3120	6	AX343070	AX343070 Sequence
24		385.4	16.1	5048	6	AX343071	AX343071 Sequence
25	c	385.4	16.1	6867	6	BD003872	BD003872 Polynucle
26		381.4	16.0	973	6	BD004035	BD004035 Polynucle
27		347	14.5	2528	6	AX343078	AX343078 Sequence
28		253.2	10.6	10312	1	AE007404	AE007404 Streptoco
29		243.4	10.2	1684	6	BD003938	BD003938 Polynucle
30		214.6	9.0	2478	6	AX088372	AX088372 Sequence
31	c	214.6	9.0	10825	1	AE006623	AE006623 Streptoco
32		213	8.9	2478	1	AB073859	AB073859 Streptoco
33	c	213	8.9	10618	1	AE010110	AE010110 Streptoco
34		211.4	8.8	1910	1	SAG290952	AJ290952 Streptoco
35		211.4	8.8	2469	6	AX088376	AX088376 Sequence
36		211.4	8.8	3501	1	AF062533	AF062533 Streptoco
37	c	205	8.6	50354	1	AE014169	AE014169 Streptoco
38		188.2	7.9	492	12	AF340223	AF340223 Synthetic
39	c	109	4.6	841	6	AR089275	AR089275 Sequence
40	c	109	4.6	841	6	AR093475	AR093475 Sequence
41	c	70.6	3.0	115758	9	AC104634	AC104634 Homo sapi
42		63.4	2.7	170522	9	AP002387	AP002387 Homo sapi
43		60.2	2.5	51939	9	AC099341	AC099341 Homo sapi
44		60	2.5	61052	2	AC117074	AC117074 Dictyoste
45		60	2.5	61052	2	AC123513	AC123513 Dictyoste

ALIGNMENTS

RESULT 1
AR120265
LOCUS AR120265 2389 bp DNA linear PAT 16-MAY-2001
DEFINITION Sequence 55 from patent US 6159469.
ACCESSION AR120265
VERSION AR120265.1 GI:14103841
KEYWORDS .
SOURCE Unknown.
ORGANISM Unknown.
Unclassified.
REFERENCE 1 (bases 1 to 2389)
AUTHORS Choi,G.H., Kunsch,C.A., Barash,S.C., Dillon,P.J., Dougherty,B., Fannon,M.R. and Rosen,C.A.
TITLE Streptococcus pneumoniae antigens and vaccines
JOURNAL Patent: US 6159469-A 55 12-DEC-2000;

```
FEATURES             Location/Qualifiers
     source          1. .2389
                     /organism="unknown"
BASE COUNT      830 a    461 c    486 g    611 t      1 others
ORIGIN
```

Query Match 100.0%; Score 2388; DB 6; Length 2389;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2389; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 TTCTTACGAGTTGGGACTGTATCAAGCTAGAACGGTTAAGGAAAATAATCGTGTTTCCTA 60
Db       1 TTCTTACGAGTTGGGACTGTATCAAGCTAGAACGGTTAAGGAAAATAATCGTGTTTCCTA 60

Qy      61 TATAGATGGAAAACAAGCGACGCAAAAAACGGAGAATTTGACTCCTGATGAGGTTAGCAA 120
Db      61 TATAGATGGAAAACAAGCGACGCAAAAAACGGAGAATTTGACTCCTGATGAGGTTAGCAA 120

Qy     121 GCGTGAAGGAATCAATGCTGAGCAAATCGTCATCAAGATAACAGACCAAGGCTATGTCAC 180
Db     121 GCGTGAAGGAATCAATGCTGAGCAAATCGTCATCAAGATAACAGACCAAGGCTATGTCAC 180

Qy     181 TTCACATGGCGACCACTATCATTATTACAATGGTAAGGTTCCTTATGACGCTATCATCAG 240
Db     181 TTCACATGGCGACCACTATCATTATTACAATGGTAAGGTTCCTTATGACGCTATCATCAG 240

Qy     241 TGAAGAATTACTCATGAAAGATCCAAACTATAAGCTAAAAGATGAGGATATTGTTAATGA 300
Db     241 TGAAGAATTACTCATGAAAGATCCAAACTATAAGCTAAAAGATGAGGATATTGTTAATGA 300

Qy     301 GGTCAAGGGTGGATATGTTATCAAGGTAGATGGAAAATACTATGTTTACCTTAAGGATGC 360
Db     301 GGTCAAGGGTGGATATGTTATCAAGGTAGATGGAAAATACTATGTTTACCTTAAGGATGC 360

Qy     361 TGCCCACGCGGATAACGTCCGTACAAAAGAGGAAATCAATCGACAAAAACAAGAGCATAG 420
Db     361 TGCCCACGCGGATAACGTCCGTACAAAAGAGGAAATCAATCGACAAAAACAAGAGCATAG 420

Qy     421 TCAACATCGTGAAGGTGGAACTCCAAGAAACGATGGTGCTGTTGCCTTGGCACGTTCGCA 480
Db     421 TCAACATCGTGAAGGTGGAACTCCAAGAAACGATGGTGCTGTTGCCTTGGCACGTTCGCA 480

Qy     481 AGGACGCTATACTACAGATGATGGTTATATCTTTAATGCTTCTGATATCATAGAGGATAC 540
Db     481 AGGACGCTATACTACAGATGATGGTTATATCTTTAATGCTTCTGATATCATAGAGGATAC 540

Qy     541 TGGTGATGCTTATATCGTTCCTCATGGAGATCATTACCATTACATTCCTAAGAATGAGTT 600
Db     541 TGGTGATGCTTATATCGTTCCTCATGGAGATCATTACCATTACATTCCTAAGAATGAGTT 600

Qy     601 ATCAGCTAGCGAGTTGGCTGCTGCAGAAGCCTTCCTATCTGGTCGAGGAAATCTGTCAAA 660
Db     601 ATCAGCTAGCGAGTTGGCTGCTGCAGAAGCCTTCCTATCTGGTCGAGGAAATCTGTCAAA 660

Qy     661 TTCAAGAACCTATCGCCGACAAAATAGCGATAACACTTCAAGAACAAACTGGGTACCTTC 720
Db     661 TTCAAGAACCTATCGCCGACAAAATAGCGATAACACTTCAAGAACAAACTGGGTACCTTC 720

Qy     721 TGTAAGCAATCCAGGAACTACAAATACTAACACAAGCAACAACAGCAACACTAACAGTCA 780
Db     721 TGTAAGCAATCCAGGAACTACAAATACTAACACAAGCAACAACAGCAACACTAACAGTCA 780

Qy     781 AGCAAGTCAAAGTAATGACATTGATAGTCTCTTGAAACAGCTCTACAAACTGCCTTTGAG 840
Db     781 AGCAAGTCAAAGTAATGACATTGATAGTCTCTTGAAACAGCTCTACAAACTGCCTTTGAG 840

Qy     841 TCAACGACATGTAGAATCTGATGGCCTTGTCTTTGATCCAGCACAAATCACAAGTCGAAC 900
Db     841 TCAACGACATGTAGAATCTGATGGCCTTGTCTTTGATCCAGCACAAATCACAAGTCGAAC 900

Qy     901 AGCTAGAGGTGTTGCAGTGCCACACGGAGATCATTACCACTTCATCCCTTACTCTCAAAT 960
Db     901 AGCTAGAGGTGTTGCAGTGCCACACGGAGATCATTACCACTTCATCCCTTACTCTCAAAT 960

Qy     961 GTCTGAATTGGAAGAACGAATCGCTCGTATTATTCCCCTTCGTTATCGTTCAAACCATTG 1020
Db     961 GTCTGAATTGGAAGAACGAATCGCTCGTATTATTCCCCTTCGTTATCGTTCAAACCATTG 1020

Qy    1021 GGTACCAGATTCAAGGCCAGAACAACCAAGTCCACAACCGACTCCGGAACCTAGTCCAGG 1080
Db    1021 GGTACCAGATTCAAGGCCAGAACAACCAAGTCCACAACCGACTCCGGAACCTAGTCCAGG 1080

Qy    1081 CCCGCAACCTGCACCAAATCTTAAAATAGACTCAAATTCTTCTTTGGTTAGTCAGCTGGT 1140
Db    1081 CCCGCAACCTGCACCAAATCTTAAAATAGACTCAAATTCTTCTTTGGTTAGTCAGCTGGT 1140

Qy    1141 ACGAAAAGTTGGGGAAGGATATGTATTCGAAGAAAAGGGCATCTCTCGTTATGTCTTTGC 1200
Db    1141 ACGAAAAGTTGGGGAAGGATATGTATTCGAAGAAAAGGGCATCTCTCGTTATGTCTTTGC 1200

Qy    1201 GAAAGATTTACCATCTGAAACTGTTAAAAATCTTGAAAGCAAGTTATCAAAACAAGAGAG 1260
Db    1201 GAAAGATTTACCATCTGAAACTGTTAAAAATCTTGAAAGCAAGTTATCAAAACAAGAGAG 1260

Qy    1261 TGTTTCACACACTTTAACTGCTAAAAAAGAAAATGTTGCTCCTCGTGACCAAGAATTTTA 1320
Db    1261 TGTTTCACACACTTTAACTGCTAAAAAAGAAAATGTTGCTCCTCGTGACCAAGAATTTTA 1320

Qy    1321 TGATAAAGCATATAATCTGTTAACTGAGGCTCATAAAGCCTTGTTTGNAAATAAGGGTCG 1380
Db    1321 TGATAAAGCATATAATCTGTTAACTGAGGCTCATAAAGCCTTGTTTGNAAATAAGGGTCG 1380

Qy    1381 TAATTCTGATTTCCAAGCCTTAGACAAATTATTAGAACGCTTGAATGATGAATCGACTAA 1440
Db    1381 TAATTCTGATTTCCAAGCCTTAGACAAATTATTAGAACGCTTGAATGATGAATCGACTAA 1440

Qy    1441 TAAAGAAAAATTGGTAGATGATTTATTGGCATTCCTAGCACCAATTACCCATCCAGAGCG 1500
Db    1441 TAAAGAAAAATTGGTAGATGATTTATTGGCATTCCTAGCACCAATTACCCATCCAGAGCG 1500

Qy    1501 ACTTGGCAAACCAAATTCTCAAATTGAGTATACTGAAGACGAAGTTCGTATTGCTCAATT 1560
Db    1501 ACTTGGCAAACCAAATTCTCAAATTGAGTATACTGAAGACGAAGTTCGTATTGCTCAATT 1560

Qy    1561 AGCTGATAAGTATACAACGTCAGATGGTTACATTTTTGATGAACATGATATAATCAGTGA 1620
Db    1561 AGCTGATAAGTATACAACGTCAGATGGTTACATTTTTGATGAACATGATATAATCAGTGA 1620

Qy    1621 TGAAGGAGATGCATATGTAACGCCTCATATGGGCCATAGTCACTGGATTGGAAAAGATAG 1680
Db    1621 TGAAGGAGATGCATATGTAACGCCTCATATGGGCCATAGTCACTGGATTGGAAAAGATAG 1680

Qy    1681 CCTTTCTGATAAGGAAAAAGTTGCAGCTCAAGCCTATACTAAAGAAAAAGGTATCCTACC 1740
Db    1681 CCTTTCTGATAAGGAAAAAGTTGCAGCTCAAGCCTATACTAAAGAAAAAGGTATCCTACC 1740

Qy    1741 TCCATCTCCAGACGCAGATGTTAAAGCAAATCCAACTGGAGATAGTGCAGCAGCTATTTA 1800
Db    1741 TCCATCTCCAGACGCAGATGTTAAAGCAAATCCAACTGGAGATAGTGCAGCAGCTATTTA 1800

Qy    1801 CAATCGTGTGAAAGGGGAAAAACGAATTCCACTCGTTCGACTTCCATATATGGTTGAGCA 1860
Db    1801 CAATCGTGTGAAAGGGGAAAAACGAATTCCACTCGTTCGACTTCCATATATGGTTGAGCA 1860

Qy    1861 TACAGTTGAGGTTAAAAACGGTAATTTGATTATTCCTCATAAGGATCATTACCATAATAT 1920
Db    1861 TACAGTTGAGGTTAAAAACGGTAATTTGATTATTCCTCATAAGGATCATTACCATAATAT 1920

Qy    1921 TAAATTTGCTTGGTTTGATGATCACACATACAAAGCTCCAAATGGCTATACCTTGGAAGA 1980
Db    1921 TAAATTTGCTTGGTTTGATGATCACACATACAAAGCTCCAAATGGCTATACCTTGGAAGA 1980

Qy    1981 TTTGTTTGCGACGATTAAGTACTACGTAGAACACCCTGACGAACGTCCACATTCTAATGA 2040
Db    1981 TTTGTTTGCGACGATTAAGTACTACGTAGAACACCCTGACGAACGTCCACATTCTAATGA 2040
```

```
Qy   2041 TGGATGGGGCAATGCCAGTGAGCATGTGTTAGGCAAGAAAGACCACAGTGAAGATCCAAA 2100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2041 TGGATGGGGCAATGCCAGTGAGCATGTGTTAGGCAAGAAAGACCACAGTGAAGATCCAAA 2100

Qy   2101 TAAGAACTTCAAAGCGGATGAAGAGCCAGTAGAGGAAACACCTGCTGAGCCAGAAGTCCC 2160
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2101 TAAGAACTTCAAAGCGGATGAAGAGCCAGTAGAGGAAACACCTGCTGAGCCAGAAGTCCC 2160

Qy   2161 TCAAGTAGAGACTGAAAAAGTAGAAGCCCAACTCAAAGAAGCAGAAGTTTTGCTTGCGAA 2220
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2161 TCAAGTAGAGACTGAAAAAGTAGAAGCCCAACTCAAAGAAGCAGAAGTTTTGCTTGCGAA 2220

Qy   2221 AGTAACGGATTCTAGTCTGAAAGCCAATGCAACAGAAACTCTAGCTGGTTTACGAAATAA 2280
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2221 AGTAACGGATTCTAGTCTGAAAGCCAATGCAACAGAAACTCTAGCTGGTTTACGAAATAA 2280

Qy   2281 TTTGACTCTTCAAATTATGGATAACAATAGTATCATGGCAGAAGCAGAAAAATTACTTGC 2340
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2281 TTTGACTCTTCAAATTATGGATAACAATAGTATCATGGCAGAAGCAGAAAAATTACTTGC 2340

Qy   2341 GTTGTTAAAAGGAAGTAATCCTTCATCTGTAAGTAAGGAAAAAATAAAC 2389
          |||||||||||||||||||||||||||||||||||||||||||||||||
Db   2341 GTTGTTAAAAGGAAGTAATCCTTCATCTGTAAGTAAGGAAAAAATAAAC 2389
```

RESULT 2
AF291695
LOCUS AF291695 2541 bp DNA linear BCT 15-MAR-2001
DEFINITION Streptococcus pneumoniae pneumococcal histidine triad A protein (phtA) gene, complete cds.
ACCESSION AF291695
VERSION AF291695.1 GI:13345012
KEYWORDS .
SOURCE Streptococcus pneumoniae.
ORGANISM Streptococcus pneumoniae
Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Streptococcus.
REFERENCE 1 (bases 1 to 2541)
AUTHORS Wizemann,T.M., Heinrichs,J.H., Adamou,J.E., Erwin,A.L., Kunsch,C., Choi,G.H., Barash,S.C., Rosen,C.A., Masure,H.R., Tuomanen,E., Gayle,A., Brewah,Y.A., Walsh,W., Barren,P., Lathigra,R., Hanson,M., Langermann,S., Johnson,S. and Koenig,S.
TITLE Use of a whole genome approach to identify vaccine molecules affording protection against Streptococcus pneumoniae infection
JOURNAL Infect. Immun. 69 (3), 1593-1598 (2001)
MEDLINE 21116976
PUBMED 11179332
REFERENCE 2 (bases 1 to 2541)
AUTHORS Choi,G.H.
TITLE Direct Submission
JOURNAL Submitted (01-AUG-2000) Molecular Biology, Human Genome Sciences, Inc., 9410 Key West Ave., Rockville, MD 20850, USA
FEATURES Location/Qualifiers
source 1. .2541
/organism="Streptococcus pneumoniae"
/strain="N4"
/db_xref="taxon:1313"
gene 91. .2541
/gene="phtA"
CDS 91. .2541
/gene="phtA"
/note="PhtA"
/codon_start=1
/transl_table=11
/product="pneumococcal histidine triad A protein"
/protein_id="AAK19155.1"
/db_xref="GI:13345013"
/translation="MKINKKYLVGSAAALILSVCSYELGLYQARTVKENNRVSYIDGK
QATQKTENLTPDEVSKREGINAEQIVIKITDQGYVTSHGDHYHYYNGKVPYDAIISEE
LLMKDPNYKLKDEDIVNEVKGGYVIKVDGKYYVYLKDAAHADNVRTKEEINRQKQEHS
QHREGGTPRNDGAVALARSQGRYTTDDGYIFNASDIIEDTGDAYIVPHGDHYHYIPKN
ELSASELAAAEAFLSGRGNLSNSRTYRRQNSDNTSRTNWVPSVSNPGTTNTNTSNNSN
TNSQASQSNDIDSLLKQLYKLPLSQRHVESDGLVFDPAQITSRTARGVAVPHGDHYHF
IPYSQMSELEERIARIIPLRYRSNHWVPDSRPEQPSPQPTPEPSPGPQPAPNLKIDSN
SSLVSQLVRKVGEGYVFEEKGISRYVFAKDLPSETVKNLESKLSKQESVSHTLTAKKE
NVAPRDQEFYDKAYNLLTEAHKALFXNKGRNSDFQALDKLLERLNDESTNKEKLVDDL
LAFLAPITHPERLGKPNSQIEYTEDEVRIAQLADKYTTSDGYIFDEHDIISDEGDAYV
TPHMGHSHWIGKDSLSDKEKVAAQAYTKEKGILPPSPDADVKANPTGDSAAAIYNRVK
GEKRIPLVRLPYMVEHTVEVKNGNLIIPHKDHYHNIKFAWFDDHTYKAPNGYTLEDLF
ATIKYYVEHPDERPHSNDGWGNASEHVLGKKDHSEDPNKNFKADEEPVEETPAEPEVP
QVETEKVEAQLKEAEVLLAKVTDSSLKANATETLAGLRNNLTLQIMDNNSIMAEAEKL
LALLKGSNPSSVSKEKIN"
BASE COUNT 888 a 476 c 516 g 660 t 1 others
ORIGIN

Query Match 100.0%; Score 2388; DB 1; Length 2541;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2389; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      1 TTCTTACGAGTTGGGACTGTATCAAGCTAGAACGGTTAAGGAAAATAATCGTGTTTCCTA 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    150 TTCTTACGAGTTGGGACTGTATCAAGCTAGAACGGTTAAGGAAAATAATCGTGTTTCCTA 209

Qy     61 TATAGATGGAAAACAAGCGACGCAAAAAACGGAGAATTTGACTCCTGATGAGGTTAGCAA 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    210 TATAGATGGAAAACAAGCGACGCAAAAAACGGAGAATTTGACTCCTGATGAGGTTAGCAA 269

Qy    121 GCGTGAAGGAATCAATGCTGAGCAAATCGTCATCAAGATAACAGACCAAGGCTATGTCAC 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    270 GCGTGAAGGAATCAATGCTGAGCAAATCGTCATCAAGATAACAGACCAAGGCTATGTCAC 329

Qy    181 TTCACATGGCGACCACTATCATTATTACAATGGTAAGGTTCCTTATGACGCTATCATCAG 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    330 TTCACATGGCGACCACTATCATTATTACAATGGTAAGGTTCCTTATGACGCTATCATCAG 389

Qy    241 TGAAGAATTACTCATGAAAGATCCAAACTATAAGCTAAAAGATGAGGATATTGTTAATGA 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    390 TGAAGAATTACTCATGAAAGATCCAAACTATAAGCTAAAAGATGAGGATATTGTTAATGA 449

Qy    301 GGTCAAGGGTGGATATGTTATCAAGGTAGATGGAAAATACTATGTTTACCTTAAGGATGC 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    450 GGTCAAGGGTGGATATGTTATCAAGGTAGATGGAAAATACTATGTTTACCTTAAGGATGC 509

Qy    361 TGCCCACGCGGATAACGTCCGTACAAAAGAGGAAATCAATCGACAAAAACAAGAGCATAG 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    510 TGCCCACGCGGATAACGTCCGTACAAAAGAGGAAATCAATCGACAAAAACAAGAGCATAG 569

Qy    421 TCAACATCGTGAAGGTGGAACTCCAAGAAACGATGGTGCTGTTGCCTTGGCACGTTCGCA 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    570 TCAACATCGTGAAGGTGGAACTCCAAGAAACGATGGTGCTGTTGCCTTGGCACGTTCGCA 629

Qy    481 AGGACGCTATACTACAGATGATGGTTATATCTTTAATGCTTCTGATATCATAGAGGATAC 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    630 AGGACGCTATACTACAGATGATGGTTATATCTTTAATGCTTCTGATATCATAGAGGATAC 689

Qy    541 TGGTGATGCTTATATCGTTCCTCATGGAGATCATTACCATTACATTCCTAAGAATGAGTT 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    690 TGGTGATGCTTATATCGTTCCTCATGGAGATCATTACCATTACATTCCTAAGAATGAGTT 749

Qy    601 ATCAGCTAGCGAGTTGGCTGCTGCAGAAGCCTTCCTATCTGGTCGAGGAAATCTGTCAAA 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    750 ATCAGCTAGCGAGTTGGCTGCTGCAGAAGCCTTCCTATCTGGTCGAGGAAATCTGTCAAA 809

Qy    661 TTCAAGAACCTATCGCCGACAAAATAGCGATAACACTTCAAGAACAAACTGGGTACCTTC 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    810 TTCAAGAACCTATCGCCGACAAAATAGCGATAACACTTCAAGAACAAACTGGGTACCTTC 869

Qy    721 TGTAAGCAATCCAGGAACTACAAATACTAACACAAGCAACAACAGCAACACTAACAGTCA 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    870 TGTAAGCAATCCAGGAACTACAAATACTAACACAAGCAACAACAGCAACACTAACAGTCA 929

Qy    781 AGCAAGTCAAAGTAATGACATTGATAGTCTCTTGAAACAGCTCTACAAACTGCCTTTGAG 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    930 AGCAAGTCAAAGTAATGACATTGATAGTCTCTTGAAACAGCTCTACAAACTGCCTTTGAG 989

Qy    841 TCAACGACATGTAGAATCTGATGGCCTTGTCTTTGATCCAGCACAAATCACAAGTCGAAC 900
```

```
Db    990 TCAACGACATGTAGAATCTGATGGCCTTGTCTTTGATCCAGCACAAATCACAAGTCGAAC 1049

Qy    901 AGCTAGAGGTGTTGCAGTGCCACACGGAGATCATTACCACTTCATCCCTTACTCTCAAAT 960
Db   1050 AGCTAGAGGTGTTGCAGTGCCACACGGAGATCATTACCACTTCATCCCTTACTCTCAAAT 1109

Qy    961 GTCTGAATTGGAAGAACGAATCGCTCGTATTATTCCCCTTCGTTATCGTTCAAACCATTG 1020
Db   1110 GTCTGAATTGGAAGAACGAATCGCTCGTATTATTCCCCTTCGTTATCGTTCAAACCATTG 1169

Qy   1021 GGTACCAGATTCAAGGCCAGAACAACCAAGTCCACAACCGACTCCGGAACCTAGTCCAGG 1080
Db   1170 GGTACCAGATTCAAGGCCAGAACAACCAAGTCCACAACCGACTCCGGAACCTAGTCCAGG 1229

Qy   1081 CCCGCAACCTGCACCAAATCTTAAAATAGACTCAAATTCTTCTTTGGTTAGTCAGCTGGT 1140
Db   1230 CCCGCAACCTGCACCAAATCTTAAAATAGACTCAAATTCTTCTTTGGTTAGTCAGCTGGT 1289

Qy   1141 ACGAAAAGTTGGGGAAGGATATGTATTCGAAGAAAAGGGCATCTCTCGTTATGTCTTTGC 1200
Db   1290 ACGAAAAGTTGGGGAAGGATATGTATTCGAAGAAAAGGGCATCTCTCGTTATGTCTTTGC 1349

Qy   1201 GAAAGATTTACCATCTGAAACTGTTAAAAATCTTGAAAGCAAGTTATCAAAACAAGAGAG 1260
Db   1350 GAAAGATTTACCATCTGAAACTGTTAAAAATCTTGAAAGCAAGTTATCAAAACAAGAGAG 1409

Qy   1261 TGTTTCACACACTTTAACTGCTAAAAAAGAAAATGTTGCTCCTCGTGACCAAGAATTTTA 1320
Db   1410 TGTTTCACACACTTTAACTGCTAAAAAAGAAAATGTTGCTCCTCGTGACCAAGAATTTTA 1469

Qy   1321 TGATAAAGCATATAATCTGTTAACTGAGGCTCATAAAGCCTTGTTTGNAAATAAGGGTCG 1380
Db   1470 TGATAAAGCATATAATCTGTTAACTGAGGCTCATAAAGCCTTGTTTGNAAATAAGGGTCG 1529

Qy   1381 TAATTCTGATTTCCAAGCCTTAGACAAATTATTAGAACGCTTGAATGATGAATCGACTAA 1440
Db   1530 TAATTCTGATTTCCAAGCCTTAGACAAATTATTAGAACGCTTGAATGATGAATCGACTAA 1589

Qy   1441 TAAAGAAAAATTGGTAGATGATTTATTGGCATTCCTAGCACCAATTACCCATCCAGAGCG 1500
Db   1590 TAAAGAAAAATTGGTAGATGATTTATTGGCATTCCTAGCACCAATTACCCATCCAGAGCG 1649

Qy   1501 ACTTGGCAAACCAAATTCTCAAATTGAGTATACTGAAGACGAAGTTCGTATTGCTCAATT 1560
Db   1650 ACTTGGCAAACCAAATTCTCAAATTGAGTATACTGAAGACGAAGTTCGTATTGCTCAATT 1709

Qy   1561 AGCTGATAAGTATACAACGTCAGATGGTTACATTTTTGATGAACATGATATAATCAGTGA 1620
Db   1710 AGCTGATAAGTATACAACGTCAGATGGTTACATTTTTGATGAACATGATATAATCAGTGA 1769

Qy   1621 TGAAGGAGATGCATATGTAACGCCTCATATGGGCCATAGTCACTGGATTGGAAAAGATAG 1680
Db   1770 TGAAGGAGATGCATATGTAACGCCTCATATGGGCCATAGTCACTGGATTGGAAAAGATAG 1829

Qy   1681 CCTTTCTGATAAGGAAAAAGTTGCAGCTCAAGCCTATACTAAAGAAAAAGGTATCCTACC 1740
Db   1830 CCTTTCTGATAAGGAAAAAGTTGCAGCTCAAGCCTATACTAAAGAAAAAGGTATCCTACC 1889

Qy   1741 TCCATCTCCAGACGCAGATGTTAAAGCAAATCCAACTGGAGATAGTGCAGCAGCTATTTA 1800
Db   1890 TCCATCTCCAGACGCAGATGTTAAAGCAAATCCAACTGGAGATAGTGCAGCAGCTATTTA 1949

Qy   1801 CAATCGTGTGAAAGGGGAAAAACGAATTCCACTCGTTCGACTTCCATATATGGTTGAGCA 1860
Db   1950 CAATCGTGTGAAAGGGGAAAAACGAATTCCACTCGTTCGACTTCCATATATGGTTGAGCA 2009

Qy   1861 TACAGTTGAGGTTAAAAACGGTAATTTGATTATTCCTCATAAGGATCATTACCATAATAT 1920
Db   2010 TACAGTTGAGGTTAAAAACGGTAATTTGATTATTCCTCATAAGGATCATTACCATAATAT 2069

Qy   1921 TAAATTTGCTTGGTTTGATGATCACACATACAAAGCTCCAAATGGCTATACCTTGGAAGA 1980
Db   2070 TAAATTTGCTTGGTTTGATGATCACACATACAAAGCTCCAAATGGCTATACCTTGGAAGA 2129

Qy   1981 TTTGTTTGCGACGATTAAGTACTACGTAGAACACCCTGACGAACGTCCACATTCTAATGA 2040
Db   2130 TTTGTTTGCGACGATTAAGTACTACGTAGAACACCCTGACGAACGTCCACATTCTAATGA 2189

Qy   2041 TGGATGGGGCAATGCCAGTGAGCATGTGTTAGGCAAGAAAGACCACAGTGAAGATCCAAA 2100
Db   2190 TGGATGGGGCAATGCCAGTGAGCATGTGTTAGGCAAGAAAGACCACAGTGAAGATCCAAA 2249

Qy   2101 TAAGAACTTCAAAGCGGATGAAGAGCCAGTAGAGGAAACACCTGCTGAGCCAGAAGTCCC 2160
Db   2250 TAAGAACTTCAAAGCGGATGAAGAGCCAGTAGAGGAAACACCTGCTGAGCCAGAAGTCCC 2309

Qy   2161 TCAAGTAGAGACTGAAAAAGTAGAAGCCCAACTCAAAGAAGCAGAAGTTTTGCTTGCGAA 2220
Db   2310 TCAAGTAGAGACTGAAAAAGTAGAAGCCCAACTCAAAGAAGCAGAAGTTTTGCTTGCGAA 2369

Qy   2221 AGTAACGGATTCTAGTCTGAAAGCCAATGCAACAGAAACTCTAGCTGGTTTACGAAATAA 2280
Db   2370 AGTAACGGATTCTAGTCTGAAAGCCAATGCAACAGAAACTCTAGCTGGTTTACGAAATAA 2429

Qy   2281 TTTGACTCTTCAAATTATGGATAACAATAGTATCATGGCAGAAGCAGAAAAATTACTTGC 2340
Db   2430 TTTGACTCTTCAAATTATGGATAACAATAGTATCATGGCAGAAGCAGAAAAATTACTTGC 2489

Qy   2341 GTTGTTAAAAGGAAGTAATCCTTCATCTGTAAGTAAGGAAAAAATAAAC 2389
Db   2490 GTTGTTAAAAGGAAGTAATCCTTCATCTGTAAGTAAGGAAAAAATAAAC 2538
```

```
RESULT 3
BD003774
LOCUS       BD003774               8195 bp    DNA     linear   PAT 31-JAN-2002
DEFINITION  Polynucleotide of Streptococcus pneumoniae and sequence.
ACCESSION   BD003774
VERSION     BD003774.1  GI:18631735
KEYWORDS    JP 2001501833-A/94.
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 8195)
  AUTHORS   Kunsch,C.A., Choi,G.H., Dillon,P.J., Rosen,C.A., Bara,S.C.,
            Fannon,M. and Dougherty,B.A.
  TITLE     Polynucleotide of Streptococcus pneumoniae and sequence
  JOURNAL   Patent: JP 2001501833-A 94 13-FEB-2001;
            HUMAN GENOME SCIENCES INC
COMMENT     OS   Unidentified
            PN   JP 2001501833-A/94
            PD   13-FEB-2001
            PF   30-OCT-1997 JP 1998520718
            PR   31-OCT-1996 US    60/029960
            PI   CHARLES A KUNSCH,GIL H CHOI,PATRICK J DILLON,CRAIG A ROSEN, PI
              STEVEN C BARASH,
            PI   MICHAEL FANNON,BRIAN A DOUGHERTY
            PC   C12N15/09,A01K67/027,C07K14/315,C07K16/12,C12N1/15,C12N1/19,
            PC   C12N1/21,
            PC   C12N5/10,C12P21/02,C12Q1/68,G06F17/30,C12N15/00,C12N5/00, PC
            G06F15/40
            CC   Strandedness: Double;
            CC   Topology: Linear;
            FH   Key             Location/Qualifiers
            FT   source          1. .8195
            FT                   /organism='Unidentified'.
FEATURES             Location/Qualifiers
     source          1. .8195
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT     2688 a   1622 c   1777 g   2105 t      3 others
ORIGIN

  Query Match           100.0%;  Score 2388;  DB 6;  Length 8195;
  Best Local Similarity 100.0%;  Pred. No. 0;
```

```
  Matches 2388;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 TTCTTACGAGTTGGGACTGTATCAAGCTAGAACGGTTAAGGAAAATAATCGTGTTTCCTA 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3053 TTCTTACGAGTTGGGACTGTATCAAGCTAGAACGGTTAAGGAAAATAATCGTGTTTCCTA 3112

Qy     61 TATAGATGGAAAACAAGCGACGCAAAAAACGGAGAATTTGACTCCTGATGAGGTTAGCAA 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3113 TATAGATGGAAAACAAGCGACGCAAAAAACGGAGAATTTGACTCCTGATGAGGTTAGCAA 3172

Qy    121 GCGTGAAGGAATCAATGCTGAGCAAATCGTCATCAAGATAACAGACCAAGGCTATGTCAC 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3173 GCGTGAAGGAATCAATGCTGAGCAAATCGTCATCAAGATAACAGACCAAGGCTATGTCAC 3232

Qy    181 TTCACATGGCGACCACTATCATTATTACAATGGTAAGGTTCCTTATGACGCTATCATCAG 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3233 TTCACATGGCGACCACTATCATTATTACAATGGTAAGGTTCCTTATGACGCTATCATCAG 3292

Qy    241 TGAAGAATTACTCATGAAAGATCCAAACTATAAGCTAAAAGATGAGGATATTGTTAATGA 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3293 TGAAGAATTACTCATGAAAGATCCAAACTATAAGCTAAAAGATGAGGATATTGTTAATGA 3352

Qy    301 GGTCAAGGGTGGATATGTTATCAAGGTAGATGGAAAATACTATGTTTACCTTAAGGATGC 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3353 GGTCAAGGGTGGATATGTTATCAAGGTAGATGGAAAATACTATGTTTACCTTAAGGATGC 3412

Qy    361 TGCCCACGCGGATAACGTCCGTACAAAAGAGGAAATCAATCGACAAAAACAAGAGCATAG 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3413 TGCCCACGCGGATAACGTCCGTACAAAAGAGGAAATCAATCGACAAAAACAAGAGCATAG 3472

Qy    421 TCAACATCGTGAAGGTGGAACTCCAAGAAACGATGGTGCTGTTGCCTTGGCACGTTCGCA 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3473 TCAACATCGTGAAGGTGGAACTCCAAGAAACGATGGTGCTGTTGCCTTGGCACGTTCGCA 3532

Qy    481 AGGACGCTATACTACAGATGATGGTTATATCTTTAATGCTTCTGATATCATAGAGGATAC 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3533 AGGACGCTATACTACAGATGATGGTTATATCTTTAATGCTTCTGATATCATAGAGGATAC 3592

Qy    541 TGGTGATGCTTATATCGTTCCTCATGGAGATCATTACCATTACATTCCTAAGAATGAGTT 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3593 TGGTGATGCTTATATCGTTCCTCATGGAGATCATTACCATTACATTCCTAAGAATGAGTT 3652

Qy    601 ATCAGCTAGCGAGTTGGCTGCTGCAGAAGCCTTCCTATCTGGTCGAGGAAATCTGTCAAA 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3653 ATCAGCTAGCGAGTTGGCTGCTGCAGAAGCCTTCCTATCTGGTCGAGGAAATCTGTCAAA 3712

Qy    661 TTCAAGAACCTATCGCCGACAAAATAGCGATAACACTTCAAGAACAAACTGGGTACCTTC 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3713 TTCAAGAACCTATCGCCGACAAAATAGCGATAACACTTCAAGAACAAACTGGGTACCTTC 3772

Qy    721 TGTAAGCAATCCAGGAACTACAAATACTAACACAAGCAACAACAGCAACACTAACAGTCA 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3773 TGTAAGCAATCCAGGAACTACAAATACTAACACAAGCAACAACAGCAACACTAACAGTCA 3832

Qy    781 AGCAAGTCAAAGTAATGACATTGATAGTCTCTTGAAACAGCTCTACAAACTGCCTTTGAG 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3833 AGCAAGTCAAAGTAATGACATTGATAGTCTCTTGAAACAGCTCTACAAACTGCCTTTGAG 3892

Qy    841 TCAACGACATGTAGAATCTGATGGCCTTGTCTTTGATCCAGCACAAATCACAAGTCGAAC 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3893 TCAACGACATGTAGAATCTGATGGCCTTGTCTTTGATCCAGCACAAATCACAAGTCGAAC 3952

Qy    901 AGCTAGAGGTGTTGCAGTGCCACACGGAGATCATTACCACTTCATCCCTTACTCTCAAAT 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3953 AGCTAGAGGTGTTGCAGTGCCACACGGAGATCATTACCACTTCATCCCTTACTCTCAAAT 4012

Qy    961 GTCTGAATTGGAAGAACGAATCGCTCGTATTATTCCCCTTCGTTATCGTTCAAACCATTG 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4013 GTCTGAATTGGAAGAACGAATCGCTCGTATTATTCCCCTTCGTTATCGTTCAAACCATTG 4072

Qy   1021 GGTACCAGATTCAAGGCCAGAACAACCAAGTCCACAACCGACTCCGGAACCTAGTCCAGG 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4073 GGTACCAGATTCAAGGCCAGAACAACCAAGTCCACAACCGACTCCGGAACCTAGTCCAGG 4132

Qy   1081 CCCGCAACCTGCACCAAATCTTAAAATAGACTCAAATTCTTCTTTGGTTAGTCAGCTGGT 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4133 CCCGCAACCTGCACCAAATCTTAAAATAGACTCAAATTCTTCTTTGGTTAGTCAGCTGGT 4192

Qy   1141 ACGAAAAGTTGGGGAAGGATATGTATTCGAAGAAAAGGGCATCTCTCGTTATGTCTTTGC 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4193 ACGAAAAGTTGGGGAAGGATATGTATTCGAAGAAAAGGGCATCTCTCGTTATGTCTTTGC 4252

Qy   1201 GAAAGATTTACCATCTGAAACTGTTAAAAATCTTGAAAGCAAGTTATCAAAACAAGAGAG 1260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4253 GAAAGATTTACCATCTGAAACTGTTAAAAATCTTGAAAGCAAGTTATCAAAACAAGAGAG 4312

Qy   1261 TGTTTCACACACTTTAACTGCTAAAAAAGAAAATGTTGCTCCTCGTGACCAAGAATTTTA 1320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4313 TGTTTCACACACTTTAACTGCTAAAAAAGAAAATGTTGCTCCTCGTGACCAAGAATTTTA 4372

Qy   1321 TGATAAAGCATATAATCTGTTAACTGAGGCTCATAAAGCCTTGTTTGNAAATAAGGGTCG 1380
          ||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db   4373 TGATAAAGCATATAATCTGTTAACTGAGGCTCATAAAGCCTTGTTTGAAAATAAGGGTCG 4432

Qy   1381 TAATTCTGATTTCCAAGCCTTAGACAAATTATTAGAACGCTTGAATGATGAATCGACTAA 1440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4433 TAATTCTGATTTCCAAGCCTTAGACAAATTATTAGAACGCTTGAATGATGAATCGACTAA 4492

Qy   1441 TAAAGAAAAATTGGTAGATGATTTATTGGCATTCCTAGCACCAATTACCCATCCAGAGCG 1500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4493 TAAAGAAAAATTGGTAGATGATTTATTGGCATTCCTAGCACCAATTACCCATCCAGAGCG 4552

Qy   1501 ACTTGGCAAACCAAATTCTCAAATTGAGTATACTGAAGACGAAGTTCGTATTGCTCAATT 1560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4553 ACTTGGCAAACCAAATTCTCAAATTGAGTATACTGAAGACGAAGTTCGTATTGCTCAATT 4612

Qy   1561 AGCTGATAAGTATACAACGTCAGATGGTTACATTTTTGATGAACATGATATAATCAGTGA 1620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4613 AGCTGATAAGTATACAACGTCAGATGGTTACATTTTTGATGAACATGATATAATCAGTGA 4672

Qy   1621 TGAAGGAGATGCATATGTAACGCCTCATATGGGCCATAGTCACTGGATTGGAAAAGATAG 1680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4673 TGAAGGAGATGCATATGTAACGCCTCATATGGGCCATAGTCACTGGATTGGAAAAGATAG 4732

Qy   1681 CCTTTCTGATAAGGAAAAAGTTGCAGCTCAAGCCTATACTAAAGAAAAAGGTATCCTACC 1740
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4733 CCTTTCTGATAAGGAAAAAGTTGCAGCTCAAGCCTATACTAAAGAAAAAGGTATCCTACC 4792

Qy   1741 TCCATCTCCAGACGCAGATGTTAAAGCAAATCCAACTGGAGATAGTGCAGCAGCTATTTA 1800
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4793 TCCATCTCCAGACGCAGATGTTAAAGCAAATCCAACTGGAGATAGTGCAGCAGCTATTTA 4852

Qy   1801 CAATCGTGTGAAAGGGGAAAAACGAATTCCACTCGTTCGACTTCCATATATGGTTGAGCA 1860
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4853 CAATCGTGTGAAAGGGGAAAAACGAATTCCACTCGTTCGACTTCCATATATGGTTGAGCA 4912

Qy   1861 TACAGTTGAGGTTAAAAACGGTAATTTGATTATTCCTCATAAGGATCATTACCATAATAT 1920
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4913 TACAGTTGAGGTTAAAAACGGTAATTTGATTATTCCTCATAAGGATCATTACCATAATAT 4972

Qy   1921 TAAATTTGCTTGGTTTGATGATCACACATACAAAGCTCCAAATGGCTATACCTTGGAAGA 1980
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4973 TAAATTTGCTTGGTTTGATGATCACACATACAAAGCTCCAAATGGCTATACCTTGGAAGA 5032

Qy   1981 TTTGTTTGCGACGATTAAGTACTACGTAGAACACCCTGACGAACGTCCACATTCTAATGA 2040
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5033 TTTGTTTGCGACGATTAAGTACTACGTAGAACACCCTGACGAACGTCCACATTCTAATGA 5092

Qy   2041 TGGATGGGGCAATGCCAGTGAGCATGTGTTAGGCAAGAAAGACCACAGTGAAGATCCAAA 2100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5093 TGGATGGGGCAATGCCAGTGAGCATGTGTTAGGCAAGAAAGACCACAGTGAAGATCCAAA 5152

Qy   2101 TAAGAACTTCAAAGCGGATGAAGAGCCAGTAGAGGAAACACCTGCTGAGCCAGAAGTCCC 2160
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5153 TAAGAACTTCAAAGCGGATGAAGAGCCAGTAGAGGAAACACCTGCTGAGCCAGAAGTCCC 5212
```

```
Qy    2161 TCAAGTAGAGACTGAAAAAGTAGAAGCCCAACTCAAAGAAGCAGAAGTTTTGCTTGCGAA 2220
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    5213 TCAAGTAGAGACTGAAAAAGTAGAAGCCCAACTCAAAGAAGCAGAAGTTTTGCTTGCGAA 5272

Qy    2221 AGTAACGGATTCTAGTCTGAAAGCCAATGCAACAGAAACTCTAGCTGGTTTACGAAATAA 2280
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    5273 AGTAACGGATTCTAGTCTGAAAGCCAATGCAACAGAAACTCTAGCTGGTTTACGAAATAA 5332

Qy    2281 TTTGACTCTTCAAATTATGGATAACAATAGTATCATGGCAGAAGCAGAAAAATTACTTGC 2340
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    5333 TTTGACTCTTCAAATTATGGATAACAATAGTATCATGGCAGAAGCAGAAAAATTACTTGC 5392

Qy    2341 GTTGTTAAAAGGAAGTAATCCTTCATCTGTAAGTAAGGAAAAAATAAAC 2389
           |||||||||||||||||||||||||||||||||||||||||||||||||
Db    5393 GTTGTTAAAAGGAAGTAATCCTTCATCTGTAAGTAAGGAAAAAATAAAC 5441


RESULT 4
AE007418/c
LOCUS       AE007418               10256 bp    DNA     linear   BCT 31-AUG-2001
DEFINITION  Streptococcus pneumoniae TIGR4  section 101 of 194 of the complete
            genome.
ACCESSION   AE007418 AE005672
VERSION     AE007418.1  GI:14972649
KEYWORDS    .
SOURCE      Streptococcus pneumoniae TIGR4.
  ORGANISM  Streptococcus pneumoniae TIGR4
            Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
            Streptococcus.
REFERENCE   1  (bases 1 to 10256)
  AUTHORS   Tettelin,H., Nelson,K.E., Paulsen,I.T., Eisen,J.A., Read,T.D.,
            Peterson,S., Heidelberg,J., DeBoy,R.T., Haft,D.H., Dodson,R.J.,
            Durkin,A.S., Gwinn,M., Kolonay,J.F., Nelson,W.C., Peterson,J.D.,
            Umayam,L.A., White,O., Salzberg,S.L., Lewis,M.R., Radune,D.,
            Holtzapple,E., Khouri,H., Wolf,A.M., Utterback,T.R., Hansen,C.L.,
            McDonald,L.A., Feldblyum,T.V., Angiuoli,S., Dickinson,T.,
            Hickey,E.K., Holt,I.E., Loftus,B.J., Yang,F., Smith,H.O.,
            Venter,J.C., Dougherty,B.A., Morrison,D.A., Hollingshead,S.K. and
            Fraser,C.M.
  TITLE     Complete genome sequence of a virulent isolate of Streptococcus
            pneumoniae
  JOURNAL   Science 293 (5529), 498-506 (2001)
  MEDLINE   21357209
   PUBMED   11463916
REFERENCE   2  (bases 1 to 10256)
  AUTHORS   Tettelin,H., Nelson,K.E., Paulsen,I.T., Eisen,J.A., Read,T.D.,
            Peterson,S., Heidelberg,J., DeBoy,R.T., Haft,D.H., Dodson,R.J.,
            Durkin,A.S., Gwinn,M., Kolonay,J.F., Nelson,W.C., Peterson,J.D.,
            Umayam,L.A., White,O., Lewis,M.R., Radune,D., Holtzapple,E.,
            Khouri,H., Wolf,A.M., Utterback,T.R., Hansen,C.L., McDonald,L.A.,
            Feldblyum,T.V., Angiuoli,S., Gesuwan,P., Hickey,E.K., Holt,I.E.,
            Loftus,B.J., Ujwal,M.L., Yang,F., Smith,H.O., Venter,J.C.,
            Dougherty,B.A., Morrison,D.A., Hollingshead,S.K. and Fraser,C.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-JUN-2001) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA
FEATURES             Location/Qualifiers
     source          1. .10256
                     /organism="Streptococcus pneumoniae TIGR4"
                     /strain="TIGR4"
                     /db_xref="taxon:170187"
     gene            complement(100. .702)
                     /gene="SP1170"
     CDS             complement(100. .702)
                     /gene="SP1170"
                     /note="identified by Glimmer2; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAK75279.1"
                     /db_xref="GI:14972650"
                     /translation="MSEVDFNEAVNYEFTSDTCQLANSIYQSLFKFFDKKNFSGDLIF
                     TWKSPSLVKEGDYIGRRDSQVDNLRVIGNIFPNYLTNRKYSLNMNRNGCMGDFPHDFF
                     DIYLDHVAKYAYEQKVNNIKEYYPLKRAILHQENALYFRFFSNFDDFLEKNYLKTIWQ
                     VSKETPFSEMDFNMFKNISEKIIFERGSKMLNDLKSNYKK"
     gene            complement(760. .1473)
                     /gene="SP1171"
     CDS             complement(760. .1473)
                     /gene="SP1171"
                     /note="identified by match to PFAM  protein family HMM
                     PF00702"
                     /codon_start=1
                     /transl_table=11
                     /product="hydrolase, haloacid dehalogenase-like family"
                     /protein_id="AAK75280.1"
                     /db_xref="GI:14972651"
                     /translation="MFYKFLLFDLDHTLLDFDAAEDVALTQLLKEEGVADIQAYKDYY
                     VPMNKALWKDLELKKISKQELVNTRFSRLFAHFGQEKDGSFLAQRYQFYLAQQGQTLS
                     GAHDLLDSLIERDYNLYAATNGITAIQTGRLAQSGLAPYFNQVFISEQLQTQKPDALF
                     YEKIGQQIAGFSKEKTLMIGDSLTADIQGGNNAGIDTIWYNPHHLENHTQAQPTYEVY
                     SYQDLLDCLDKNILEKITF"
     gene            complement(1724. .1843)
                     /gene="SP1172"
     CDS             complement(1724. .1843)
                     /gene="SP1172"
                     /note="identified by Glimmer2; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAK75281.1"
                     /db_xref="GI:14972652"
                     /translation="MLYAVPFYFNRSETIVFLNCESIKTDCDGAILALETFKN"
     gene            complement(1933. .2307)
                     /gene="SP1173"
     CDS             complement(1933. .2307)
                     /gene="SP1173"
                     /note="identified by Glimmer2; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="conserved hypothetical protein"
                     /protein_id="AAK75282.1"
                     /db_xref="GI:14972653"
                     /translation="MESIGLVIVSHSKHIAEGVVELISKVAKDVPITYVRGTEGGGIG
                     TSFEQVDRVVSENPADTLLAFFDLGSAKMNLKMVTDFSDKSIIINRVPIVEGAYNAAA
                     LLQAGAELSVIQTQLAELEINK"
     gene            complement(2729. .5188)
                     /gene="SP1174"
     CDS             complement(2729. .5188)
                     /gene="SP1174"
                     /note="similar to GP:4249624; identified by sequence
                     similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="conserved domain protein"
                     /protein_id="AAK75283.1"
                     /db_xref="GI:14972654"
                     /translation="MKINKKYLAGSVAVLALSVCSYELGRYQAGQDKKESNRVAYIDG
                     DQAGQKAENLTPDEVSKREGINAEQIVIKITDQGYVTSHGDHYHYYNGKVPYDAIISE
                     ELLMKDPNYQLKDSDIVNEIKGGYVIKVNGKYYVYLKDAAHADNIRTKEEIKRQKQER
                     SHNHNSRADNAVAAARAQGRYTTDDGYIFNASDIIEDTGDAYIVPHGDHYHYIPKNEL
                     SASELAAAEAYWNGKQGSRPSSSSSYNANPAQPRLSENHNLTVTPTYHQNQGENISSL
                     LRELYAKPLSERHVESDGLIFDPAQITSRTARGVAVPHGNHYHFIPYEQMSELEKRIA
                     RIIPLRYRSNHWVPDSRPEEPSPQPTPEPSPSPQPAPSNPIDEKLVKEAVRKVGDGYV
                     FEENGVSRYIPAKDLSAETAAGIDSKLAKQESLSHKLGTKKTDLPSSDREFYNKAYDL
                     LARIHQDLLDNKGRQVDFEALDNLLERLKDVSSDKVKLVEDILAFLAPIRHPERLGKP
                     NAQITYTDDEIQVAKLAGKYTTEDGYIFDPRDITSDEGDAYVTPHMTHSHWIKKDSLS
                     EAERAAAQAYAKEKGLTPPSTDHQDSGNTEAKGAEAIYNRVKAAKKVPLDRMPYNLQY
                     TVEVKNGSLIIPHYDHYHNIKFEWFDEGLYEAPKGYTLEDLLATVKYYVEHPNERPHS
                     DNGFGNASDHVQRNKNGQADTNQTEKPSEEKPQTEKPEEETPREEKPQSEKPESPKPT
                     EEPEESPEESEEPQVETEKVEEKLREAEDLLGKIQDPIIKSNAKETLTGLKNNLLFGT
                     QDNNTIMAEAEKLLALLKESK"
     gene            complement(5346. .7754)
                     /gene="SP1175"
     CDS             complement(5346. .7754)
                     /gene="SP1175"
                     /note="similar to GP:4249624; identified by sequence
```

```
                     similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="conserved domain protein"
                     /protein_id="AAK75284.1"
                     /db_xref="GI:14972655"
                     /translation="MILSVCSYELGLYQARTVKENNRVSYIDGKQATQKTENLTPDEV
                     SKREGINAEQIVIKITDQGYVTSHGDHYHYYNGKVPYDAIISEELLMKDPNYKLKDED
                     IVNEVKGGYVIKVDGKYYVYLKDAAHADNVRTKEEINRQKQEHSQHREGGTPRNDGAV
                     ALARSQGRYTTDDGYIFNASDIIEDTGDAYIVPHGDHYHYIPKNELSASELAAAEAFL
                     SGRGNLSNSRTYRRQNSDNTSRTNWVPSVSNPGTTNTNTSNNSNTNSQASQSNDIDSL
                     LKQLYKLPLSQRHVESDGLVFDPAQITSRTARGVAVPHGDHYHFIPYSQMSELEERIA
                     RIIPLRYRSNHWVPDSRPEQPSPQPTPEPSPGPQPAPNLKIDSNSSLVSQLVRKVGEG
                     YVFEEKGISRYVFAKDLPSETVKNLESKLSKQESVSHTLTAKKENVAPRDQEFYDKAY
                     NLLTEAHKALFENKGRNSDFQALDKLLERLNDESTNKEKLVDDLLAFLAPITHPERLG
                     KPNSQIEYTEDEVRIAQLADKYTTSDGYIFDEHDIISDEGDAYVTPHMGHSHWIGKDS
                     LSDKEKVAAQAYTKEKGILPPSPDADVKANPTGDSAAAIYNRVKGEKRIPLVRLPYMV
                     EHTVEVKNGNLIIPHKDHYHNIKFAWFDDHTYKAPNGYTLEDLFATIKYYVEHPDERP
                     HSNDGWGNASEHVLGKKDHSEDPNKNFKADEEPVEETPAEPEVPQVETEKVEAQLKEA
                     EVLLAKVTDSSLKANATETLAGLRNNLTLQIMDNNSIMAEAEKLLALLKGSNPSSVSK
                     EKIN"
     gene            complement(8049. .9782)
                     /gene="SP1176"
     CDS             complement(8049. .9782)
                     /gene="SP1176"
                     /codon_start=1
                     /transl_table=11
                     /product="phosphoenolpyruvate-protein phosphotransferase"
                     /protein_id="AAK75285.1"
                     /db_xref="GI:14972656"
                     /translation="MTEMLKGIAASDGVAVAKAYLLVQPDLSFETITVEDTNAEEARL
                     DAALQASQDELSVIREKAVGTLGEEAAQVFDAHLMVLADPEMISQIKETIRAKKVNAE
                     AGLKEVTDMFITIFEGMEDNPYMQERAADIRDVTKRVLANLLGKKLPNPASINEEVIV
                     IAHDLTPSDTAQLDKNFVKAFVTNIGGRTSHSAIMARTLEIAAVLGTNNITEIVKDGD
                     ILAVNGITGEVIINPTDEQAAEFKAAGEAYAKQKAEWALLKDAQTVTADGKHFELAAN
                     IGTPKDVEGVNNNGAEAVGLYRTEFLYMDSQDFPTEDEQYEAYKAVLEGMNGKPVVVR
                     TMDIGGDKELPYFDMPHEMNPFLGFRALRISISETGDAMFRTQIRALLRASVHGQLRI
                     MFPMVALLKEFRAAKAVFDEEKANLLAEGVAVADNIQVGIMIEIPAAAMLADQFAKEV
                     DFFSIGTNDLIQYTMAADRMNEQVSYLYQPYNPSILRLINNVIKAAHAEGKWAGMCGE
                     MAGDQQAVPLLVGMGLDEFSMSATSVLRTRSLMKKLDTAKMEEYANRALTECSTMEEV
                     LELQKEYVNFD"
     gene            complement(9788. .10051)
                     /gene="SP1177"
     CDS             complement(9788. .10051)
                     /gene="SP1177"
                     /codon_start=1
                     /transl_table=11
                     /product="phosphocarrier protein HPr"
                     /protein_id="AAK75286.1"
                     /db_xref="GI:14972657"
                     /translation="MASKDFHVVAETGIHARPATLLVQTASKFASDITLEYKGKSVNL
                     KSIMGVMSLGVGQGADVTISAEGADADDAIAAISETMEKEGLA"
BASE COUNT     2842 a    2089 c    1850 g    3475 t
ORIGIN

  Query Match           100.0%;  Score 2388;  DB 1;  Length 10256;
  Best Local Similarity  100.0%;  Pred. No. 0;
  Matches 2388;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 TTCTTACGAGTTGGGACTGTATCAAGCTAGAACGGTTAAGGAAAATAATCGTGTTTCCTA 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7737 TTCTTACGAGTTGGGACTGTATCAAGCTAGAACGGTTAAGGAAAATAATCGTGTTTCCTA 7678

Qy     61 TATAGATGGAAAACAAGCGACGCAAAAAACGGAGAATTTGACTCCTGATGAGGTTAGCAA 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7677 TATAGATGGAAAACAAGCGACGCAAAAAACGGAGAATTTGACTCCTGATGAGGTTAGCAA 7618

Qy    121 GCGTGAAGGAATCAATGCTGAGCAAATCGTCATCAAGATAACAGACCAAGGCTATGTCAC 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7617 GCGTGAAGGAATCAATGCTGAGCAAATCGTCATCAAGATAACAGACCAAGGCTATGTCAC 7558

Qy    181 TTCACATGGCGACCACTATCATTATTACAATGGTAAGGTTCCTTATGACGCTATCATCAG 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7557 TTCACATGGCGACCACTATCATTATTACAATGGTAAGGTTCCTTATGACGCTATCATCAG 7498

Qy    241 TGAAGAATTACTCATGAAAGATCCAAACTATAAGCTAAAAGATGAGGATATTGTTAATGA 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7497 TGAAGAATTACTCATGAAAGATCCAAACTATAAGCTAAAAGATGAGGATATTGTTAATGA 7438

Qy    301 GGTCAAGGGTGGATATGTTATCAAGGTAGATGGAAAATACTATGTTTACCTTAAGGATGC 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7437 GGTCAAGGGTGGATATGTTATCAAGGTAGATGGAAAATACTATGTTTACCTTAAGGATGC 7378

Qy    361 TGCCCACGCGGATAACGTCCGTACAAAAGAGGAAATCAATCGACAAAAACAAGAGCATAG 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7377 TGCCCACGCGGATAACGTCCGTACAAAAGAGGAAATCAATCGACAAAAACAAGAGCATAG 7318

Qy    421 TCAACATCGTGAAGGTGGAACTCCAAGAAACGATGGTGCTGTTGCCTTGGCACGTTCGCA 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7317 TCAACATCGTGAAGGTGGAACTCCAAGAAACGATGGTGCTGTTGCCTTGGCACGTTCGCA 7258

Qy    481 AGGACGCTATACTACAGATGATGGTTATATCTTTAATGCTTCTGATATCATAGAGGATAC 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7257 AGGACGCTATACTACAGATGATGGTTATATCTTTAATGCTTCTGATATCATAGAGGATAC 7198

Qy    541 TGGTGATGCTTATATCGTTCCTCATGGAGATCATTACCATTACATTCCTAAGAATGAGTT 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7197 TGGTGATGCTTATATCGTTCCTCATGGAGATCATTACCATTACATTCCTAAGAATGAGTT 7138

Qy    601 ATCAGCTAGCGAGTTGGCTGCTGCAGAAGCCTTCCTATCTGGTCGAGGAAATCTGTCAAA 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7137 ATCAGCTAGCGAGTTGGCTGCTGCAGAAGCCTTCCTATCTGGTCGAGGAAATCTGTCAAA 7078

Qy    661 TTCAAGAACCTATCGCCGACAAAATAGCGATAACACTTCAAGAACAAACTGGGTACCTTC 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7077 TTCAAGAACCTATCGCCGACAAAATAGCGATAACACTTCAAGAACAAACTGGGTACCTTC 7018

Qy    721 TGTAAGCAATCCAGGAACTACAAATACTAACACAAGCAACAACAGCAACACTAACAGTCA 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7017 TGTAAGCAATCCAGGAACTACAAATACTAACACAAGCAACAACAGCAACACTAACAGTCA 6958

Qy    781 AGCAAGTCAAAGTAATGACATTGATAGTCTCTTGAAACAGCTCTACAAACTGCCTTTGAG 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6957 AGCAAGTCAAAGTAATGACATTGATAGTCTCTTGAAACAGCTCTACAAACTGCCTTTGAG 6898

Qy    841 TCAACGACATGTAGAATCTGATGGCCTTGTCTTTGATCCAGCACAAATCACAAGTCGAAC 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6897 TCAACGACATGTAGAATCTGATGGCCTTGTCTTTGATCCAGCACAAATCACAAGTCGAAC 6838

Qy    901 AGCTAGAGGTGTTGCAGTGCCACACGGAGATCATTACCACTTCATCCCTTACTCTCAAAT 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6837 AGCTAGAGGTGTTGCAGTGCCACACGGAGATCATTACCACTTCATCCCTTACTCTCAAAT 6778

Qy    961 GTCTGAATTGGAAGAACGAATCGCTCGTATTATTCCCCTTCGTTATCGTTCAAACCATTG 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6777 GTCTGAATTGGAAGAACGAATCGCTCGTATTATTCCCCTTCGTTATCGTTCAAACCATTG 6718

Qy   1021 GGTACCAGATTCAAGGCCAGAACAACCAAGTCCACAACCGACTCCGGAACCTAGTCCAGG 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6717 GGTACCAGATTCAAGGCCAGAACAACCAAGTCCACAACCGACTCCGGAACCTAGTCCAGG 6658

Qy   1081 CCCGCAACCTGCACCAAATCTTAAAATAGACTCAAATTCTTCTTTGGTTAGTCAGCTGGT 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6657 CCCGCAACCTGCACCAAATCTTAAAATAGACTCAAATTCTTCTTTGGTTAGTCAGCTGGT 6598

Qy   1141 ACGAAAAGTTGGGGAAGGATATGTATTCGAAGAAAAGGGCATCTCTCGTTATGTCTTTGC 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6597 ACGAAAAGTTGGGGAAGGATATGTATTCGAAGAAAAGGGCATCTCTCGTTATGTCTTTGC 6538

Qy   1201 GAAAGATTTACCATCTGAAACTGTTAAAAATCTTGAAAGCAAGTTATCAAAACAAGAGAG 1260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6537 GAAAGATTTACCATCTGAAACTGTTAAAAATCTTGAAAGCAAGTTATCAAAACAAGAGAG 6478

Qy   1261 TGTTTCACACACTTTAACTGCTAAAAAAGAAAATGTTGCTCCTCGTGACCAAGAATTTTA 1320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6477 TGTTTCACACACTTTAACTGCTAAAAAAGAAAATGTTGCTCCTCGTGACCAAGAATTTTA 6418
```

```
Qy   1321 TGATAAAGCATATAATCTGTTAACTGAGGCTCATAAAGCCTTGTTTGNAAATAAGGGTCG 1380
          ||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db   6417 TGATAAAGCATATAATCTGTTAACTGAGGCTCATAAAGCCTTGTTTGAAAATAAGGGTCG 6358

Qy   1381 TAATTCTGATTTCCAAGCCTTAGACAAATTATTAGAACGCTTGAATGATGAATCGACTAA 1440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6357 TAATTCTGATTTCCAAGCCTTAGACAAATTATTAGAACGCTTGAATGATGAATCGACTAA 6298

Qy   1441 TAAAGAAAAATTGGTAGATGATTTATTGGCATTCCTAGCACCAATTACCCATCCAGAGCG 1500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6297 TAAAGAAAAATTGGTAGATGATTTATTGGCATTCCTAGCACCAATTACCCATCCAGAGCG 6238

Qy   1501 ACTTGGCAAACCAAATTCTCAAATTGAGTATACTGAAGACGAAGTTCGTATTGCTCAATT 1560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6237 ACTTGGCAAACCAAATTCTCAAATTGAGTATACTGAAGACGAAGTTCGTATTGCTCAATT 6178

Qy   1561 AGCTGATAAGTATACAACGTCAGATGGTTACATTTTTGATGAACATGATATAATCAGTGA 1620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6177 AGCTGATAAGTATACAACGTCAGATGGTTACATTTTTGATGAACATGATATAATCAGTGA 6118

Qy   1621 TGAAGGAGATGCATATGTAACGCCTCATATGGGCCATAGTCACTGGATTGGAAAAGATAG 1680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6117 TGAAGGAGATGCATATGTAACGCCTCATATGGGCCATAGTCACTGGATTGGAAAAGATAG 6058

Qy   1681 CCTTTCTGATAAGGAAAAAGTTGCAGCTCAAGCCTATACTAAAGAAAAAGGTATCCTACC 1740
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6057 CCTTTCTGATAAGGAAAAAGTTGCAGCTCAAGCCTATACTAAAGAAAAAGGTATCCTACC 5998

Qy   1741 TCCATCTCCAGACGCAGATGTTAAAGCAAATCCAACTGGAGATAGTGCAGCAGCTATTTA 1800
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5997 TCCATCTCCAGACGCAGATGTTAAAGCAAATCCAACTGGAGATAGTGCAGCAGCTATTTA 5938

Qy   1801 CAATCGTGTGAAAGGGGAAAAACGAATTCCACTCGTTCGACTTCCATATATGGTTGAGCA 1860
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5937 CAATCGTGTGAAAGGGGAAAAACGAATTCCACTCGTTCGACTTCCATATATGGTTGAGCA 5878

Qy   1861 TACAGTTGAGGTTAAAAACGGTAATTTGATTATTCCTCATAAGGATCATTACCATAATAT 1920
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5877 TACAGTTGAGGTTAAAAACGGTAATTTGATTATTCCTCATAAGGATCATTACCATAATAT 5818

Qy   1921 TAAATTTGCTTGGTTTGATGATCACACATACAAAGCTCCAAATGGCTATACCTTGGAAGA 1980
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5817 TAAATTTGCTTGGTTTGATGATCACACATACAAAGCTCCAAATGGCTATACCTTGGAAGA 5758

Qy   1981 TTTGTTTGCGACGATTAAGTACTACGTAGAACACCCTGACGAACGTCCACATTCTAATGA 2040
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5757 TTTGTTTGCGACGATTAAGTACTACGTAGAACACCCTGACGAACGTCCACATTCTAATGA 5698

Qy   2041 TGGATGGGGCAATGCCAGTGAGCATGTGTTAGGCAAGAAAGACCACAGTGAAGATCCAAA 2100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5697 TGGATGGGGCAATGCCAGTGAGCATGTGTTAGGCAAGAAAGACCACAGTGAAGATCCAAA 5638

Qy   2101 TAAGAACTTCAAAGCGGATGAAGAGCCAGTAGAGGAAACACCTGCTGAGCCAGAAGTCCC 2160
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5637 TAAGAACTTCAAAGCGGATGAAGAGCCAGTAGAGGAAACACCTGCTGAGCCAGAAGTCCC 5578

Qy   2161 TCAAGTAGAGACTGAAAAAGTAGAAGCCCAACTCAAAGAAGCAGAAGTTTTGCTTGCGAA 2220
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5577 TCAAGTAGAGACTGAAAAAGTAGAAGCCCAACTCAAAGAAGCAGAAGTTTTGCTTGCGAA 5518

Qy   2221 AGTAACGGATTCTAGTCTGAAAGCCAATGCAACAGAAACTCTAGCTGGTTTACGAAATAA 2280
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5517 AGTAACGGATTCTAGTCTGAAAGCCAATGCAACAGAAACTCTAGCTGGTTTACGAAATAA 5458

Qy   2281 TTTGACTCTTCAAATTATGGATAACAATAGTATCATGGCAGAAGCAGAAAAATTACTTGC 2340
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5457 TTTGACTCTTCAAATTATGGATAACAATAGTATCATGGCAGAAGCAGAAAAATTACTTGC 5398

Qy   2341 GTTGTTAAAAGGAAGTAATCCTTCATCTGTAAGTAAGGAAAAAATAAAC 2389
          |||||||||||||||||||||||||||||||||||||||||||||||||
Db   5397 GTTGTTAAAAGGAAGTAATCCTTCATCTGTAAGTAAGGAAAAAATAAAC 5349
```

```
RESULT 5
AE008479/c
LOCUS       AE008479               10320 bp    DNA     linear   BCT 13-SEP-2001
DEFINITION  Streptococcus pneumoniae R6 section 95 of 184 of the complete
            genome.
ACCESSION   AE008479 AE007317
VERSION     AE008479.1  GI:15458677
KEYWORDS    .
SOURCE      Streptococcus pneumoniae R6.
  ORGANISM  Streptococcus pneumoniae R6
            Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
            Streptococcus.
REFERENCE   1  (bases 1 to 10320)
  AUTHORS   Hoskins,J.A., Alborn,W. Jr., Arnold,J., Blaszczak,L., Burgett,S.,
            DeHoff,B.S., Estrem,S., Fritz,L., Fu,D.-J., Fuller,W., Geringer,C.,
            Gilmour,R., Glass,J.S., Khoja,H., Kraft,A., LaGace,R.,
            LeBlanc,D.J., Lee,L.N., Lefkowitz,E.J., Lu,J., Matsushima,P.,
            McAhren,S., McHenney,M., McLeaster,K., Mundy,C., Nicas,T.I.,
            Norris,F.H., O'Gara,M., Peery,R., Robertson,G.T., Rockey,P.,
            Sun,P.-M., Winkler,M.E., Yang,Y., Young-Bellido,M., Zhao,G.,
            Zook,C., Baltz,R.H., Jaskunas,S.R., Rosteck,P.R. Jr., Skatrud,P.L.
            and Glass,J.I.
  TITLE     Genome of the bacterium Streptococcus pneumoniae strain R6
  JOURNAL   J. Bacteriol. 183 (19), 5709-5717 (2001)
  MEDLINE   21429245
   PUBMED   11544234
REFERENCE   2  (bases 1 to 10320)
  AUTHORS   Hoskins,J.A., Alborn,W. Jr., Arnold,J., Blaszczak,L., Burgett,S.,
            DeHoff,B.S., Estrem,S., Fritz,L., Fu,D.-J., Fuller,W., Geringer,C.,
            Gilmour,R., Glass,J.S., Hann,A., Khoja,H., Kraft,A., LaGace,R.,
            LeBlanc,D.J., Lee,L.N., Lefkowitz,E.J., Lu,J., Matsushima,P.,
            McAhren,S., McHenney,M., McLeaster,K., Mundy,C., Nicas,T.I.,
            Norris,F.H., O'Gara,M., Peery,R., Robertson,G.T., Rockey,P.,
            Sun,P.-M., Winkler,M.E., Yang,Y., Young-Bellido,M., Zhao,G.,
            Zook,C., Baltz,R.H., Jaskunas,S.R., Rosteck,P.R. Jr., Skatrud,P.L.
            and Glass,J.I.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-JUL-2001) Infectious Diseases Research, Eli Lilly and
            Company, Lilly Research Labs, Indianapolis, IN 46285-0438, USA
FEATURES             Location/Qualifiers
     source          1. .10320
                     /organism="Streptococcus pneumoniae R6"
                     /strain="R6"
                     /db_xref="taxon:171101"
     gene            complement(91. .738)
                     /gene="spr1056"
     CDS             complement(91. .738)
                     /gene="spr1056"
                     /codon_start=1
                     /transl_table=11
                     /label=spr1056
                     /product="Hypothetical protein"
                     /protein_id="AAK99860.1"
                     /db_xref="GI:15458678"
                     /translation="MTTKKANITIEEYIEMSEVDFNEAVNYEFTSDTCQLANSIYQSL
                     FKFFDKKNFSGDLIFTWKSPSLVKEGDYIGRRDSQVDNLRVIGNIFPNYLTNRKYSLN
                     MNRNGCMGDFPHDFFDIYLDHVAKYAYEQKVNNIKEYYPLKRAILHQENALYFRLFSN
                     FDDFLEKNYLKTIWQVSKETPFSEMDFNMFKNISEKIIFERGSKMLNDLKSNYKK"
     gene            complement(752. .1465)
                     /gene="spr1057"
     CDS             complement(752. .1465)
                     /gene="spr1057"
                     /codon_start=1
                     /transl_table=11
                     /label=spr1057
                     /product="Conserved hypothetical protein"
                     /protein_id="AAK99861.1"
                     /db_xref="GI:15458679"
                     /translation="MFYKFLLFDLDHTLLDFDAAEDVALTQLLKEEGVADIQAYKDYY
                     VPMNKALWKDLELKKISKQELVNTRFSRLFSHFGQEKDGSFLAQRYQFYLAQQGQTLS
                     GAHDLLDSLIERDYDLYAATNGITAIQTGRLAQSGLVPYFNQVFISEQLQTQKPDALF
                     YEKIGQQIAGFSKEKTLMIGDSLTADIQGGNNAGIDTIWYNPHHLENHTQAQPTYEVY
```

```
                SYQDLLDCLDKNILEKITF"
gene            complement(1923. .2081)
                /gene="spr1058"
CDS             complement(1923. .2081)
                /gene="spr1058"
                /codon_start=1
                /transl_table=11
                /label=spr1058
                /product="Hypothetical protein"
                /protein_id="AAK99862.1"
                /db_xref="GI:15458680"
                /translation="MNLKMVTDFSDKSIIINRVPIVEGAYNAAALLQAGAELSVIQTQ
                LAELEINK"
gene            complement(2151. .2306)
                /gene="spr1059"
CDS             complement(2151. .2306)
                /gene="spr1059"
                /codon_start=1
                /transl_table=11
                /label=spr1059
                /product="Hypothetical protein"
                /protein_id="AAK99863.1"
                /db_xref="GI:15458681"
                /translation="MMESIGIGLVIVSHSKHIAEGVVELISKVAKDVPITYVGGTEGG
                GIGTSFD"
repeat_region   complement(2542. .2613)
                /rpt_family="RUP element"
                /rpt_type=dispersed
                /rpt_unit=Rup_A
gene            complement(2725. .5292)
                /gene="phpA"
                /note="spr1060"
CDS             complement(2725. .5292)
                /gene="phpA"
                /codon_start=1
                /transl_table=11
                /label=spr1060
                /product="Histidine Motif-Containing protein"
                /protein_id="AAK99864.1"
                /db_xref="GI:15458682"
                /translation="MNQIYLRKEERMKINKKYLAGSVATLVLSVCAYELGLHQAQTVK
                ENNRVSYIDGKQATQKTENLTPDEVSKREGINAEQIVIKITDQGYVTSHGDHYHYYNG
                KVPYDAIISEELLMKDPNYQLKDEDIISEIKGGYVIKVDGKYYVYLKDAAHADNVRTK
                EEINRQKQEHSQHREGGTPRNDGAVALARSQGRYTTDDGYIFNASDIIEDTGDAYIVP
                HGDHYHYIPKNELSASELAAAKAFLSGRGNLSNSRTYRRQNSDNTSRTNWVPSVSNPG
                TTNTNTSNNSNTNSQASQSNDIDSLLKQLYKLPLSQRHVESDGLIFDPAQITSRTANG
                VAVPHGDHYHFIPYSQLSPLEEKLARIIPLRYRSNHWVPDSRPEQPSPQSTPEPSPSP
                QPAPNPQPAPSNPIDEKLVKEAVRKVGDGYVFEENGVPRYIPAKDLSAETAAGIDSKL
                AKQESLSHKLGAKKTDLPSSDREFYNKAYDLLARIHQDLLDNKGRQVDFEALDNLLER
                LKDVSSDKVKLVDDILAFLAPIRHPERLGKPNAQITYTDDEIQVAKLAGKYTTEDGYI
                FDPRDITSDEGDAYVTPHMTHSHWIKKDSLSEAERAAAQAYAKEKGLTPPSTDHQDSG
                NTEAKGAEAIYNRVKAAKKVPLDRMPYNLQYTVEVKNGSLIIPHYDHYHNIKFEWFDE
                GLYEAPKGYSLEDLLATVKYYVEHPNERPHSDNGFGNASDHVQRNKNGQADTNQTEKP
                NEEKPQTEKPEEETPREEKPQSEKPESPKPTEEPEEESPEESPEESEEPQVETEKVKE
                KLREAEDLLGKIQNPIIKSNAKETLTGLKNNLLFGTQDNNTIMAEAEKLLALLKESK"
gene            complement(5417. .7903)
                /gene="phtA"
                /note="spr1061"
CDS             complement(5417. .7903)
                /gene="phtA"
                /codon_start=1
                /transl_table=11
                /label=spr1061
                /product="Pneumococcal histidine triad protein A
                precursor"
                /protein_id="AAK99865.1"
                /db_xref="GI:15458683"
                /translation="MQLEISNRKRVSMKINKKYLVGSAAALILSVCSYELGLYQARTV
                KENNRVSYIDGKQATQKTENLTPDEVSKREGINAEQIVIKITDQGYVTSHGDHYHYYN
                GKVPYDAIFSEELLMKDPNYKLKDEDIVNEVKGGYVIKVDGKYYVYLKDAAHADNVRT
                KEEINRQKQEHSQHREGGTPRNDGAVALARSQGRYTTDDGYIFNASDIIEDTGDAYIV
                PHGDHYHYIPKNELSASELAAAEAFLSGRGNLSNSRTYRRQNSDNTSRTNWVPSVSNP
                GTTNTNTSNNSNTNSQASQSNDIDSLLKQLYKLPLSQRHVESDGLVFDPAQITSRTAR
                GVAVPHGDHYHFIPYSQMSELEERIARIIPLRYRSNHWVPDSRPEQPSPQPTPEPSPG
                PQPAPNLKIDSNSSLVSQLVRKVGEGYVFEEKGISRYVFAKDLPSETVKNLESKLSKQ
                ESVSHTLTAKKENVAPRDQEFYDKAYNLLTEAHKALFENKGRNSDFQALDKLLERLND
                ESTNKEKLVDDLLAFLAPITHPERLGKPNSQIEYTEDEVRIAQLADKYTTSDGYIFDE
                HDIISDEGDAYVTPHMGHSHWIGKDSLSDKEKVAAQAYTKEKGILPPSPDADVKANPT
                GDSAAAIYNRVKGEKRIPLVRLPYMVEHTVEVKNGNLIIPHKDHYHNIKFAWFDDHTY
                KAPNGYTLEDLFATIKYYVEHPDERPHSNDGWGNASEHVLGKKDHSEDPNKNFKADEE
                PVEETPAEPEVPQVETEKVEAQLKEAEVLLAKVTDSSLKANATETLAGLRNNLTLQIM
                DNNSIMAEAEKLLALLKGSNPSSVSKEKIN"
gene            complement(8122. .9855)
                /gene="ptsI"
                /note="spr1062"
CDS             complement(8122. .9855)
                /gene="ptsI"
                /EC_number="2.7.3.9"
                /codon_start=1
                /transl_table=11
                /label=spr1062
                /product="PTS enzyme I"
                /protein_id="AAK99866.1"
                /db_xref="GI:15458684"
                /translation="MTEMLKGIAASDGVAVAKAYLLVQPDLSFETITVEDTNAEEARL
                DAALQASQDELSVIREKAVGTLGEEAAQVFDAHLMVLADPEMISQIKETIRAKKVNAE
                AGLKEVTDMFITIFEGMEDNPYMQERAADIRDVTKRVLANLLGKKLPNPASINEEVIV
                IAHDLTPSDTAQLDKNFVKAFVTNIGGRTSHSAIMARTLEIAAVLGTNNITEIVKDGD
                ILAVNGITGEVIINPTDEQAAEFKAAGEAYAKQKAEWALLKDAQTVTADGKHFELAAN
                IGTPKDVEGVNNNGAEAVGLYRTEFLYMDSQDFPTEDEQYEAYKAVLEGMNGKPVVVR
                TMDIGGDKELPYFDMPHEMNPFLGFRALRISISETGDAMFRTQIRALLRASVHGQLRI
                MFPMVALLKEFRAAKAVFDEEKANLLAEGVAVADNIQVGIMIEIPAAAMLADQFAKEV
                DFFSIGTNDLIQYTMAADRMNEQVSYLYQPYNPSILRLINNVIKAAHAEGKWAGMCGE
                MAGDQQAVPLLVGMGLDEFSMSATSVLRTRSLMKKLDTAKMEEYANRALTECSTMEEV
                LELQKEYVNFD"
gene            complement(9861. .10124)
                /gene="ptsH"
                /note="spr1063"
CDS             complement(9861. .10124)
                /gene="ptsH"
                /codon_start=1
                /transl_table=11
                /label=spr1063
                /product="Histidine-containing phosphocarrier protein of
                the PTS"
                /protein_id="AAK99867.1"
                /db_xref="GI:15458685"
                /translation="MASKDFHVVAETGIHARPATLLVQTASKFASDITLEYKGKSVNL
                KSIMGVMSLGVGQGADVTISAEGADADDAIAAISETMEKEGLA"
BASE COUNT      2851 a   2091 c   1862 g   3516 t
ORIGIN
```

```
  Query Match             99.0%;  Score 2365.6;  DB 1;  Length 10320;
  Best Local Similarity   99.4%;  Pred. No. 0;
  Matches 2374;  Conservative    0;  Mismatches   15;  Indels    0;  Gaps    0;

Qy       1 TTCTTACGAGTTGGGACTGTATCAAGCTAGAACGGTTAAGGAAAATAATCGTGTTTCCTA 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    7808 TTCTTACGAGTTGGGACTGTATCAAGCTAGAACGGTTAAGGAAAATAATCGTGTTTCCTA 7749

Qy      61 TATAGATGGAAAACAAGCGACGCAAAAAACGGAGAATTTGACTCCTGATGAGGTTAGCAA 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    7748 TATAGATGGAAAACAAGCGACGCAAAAAACGGAGAATTTGACTCCTGATGAGGTTAGCAA 7689

Qy     121 GCGTGAAGGAATCAATGCTGAGCAAATCGTCATCAAGATAACAGACCAAGGCTATGTCAC 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    7688 GCGTGAAGGAATCAATGCTGAGCAAATCGTCATCAAGATAACAGACCAAGGCTATGTCAC 7629

Qy     181 TTCACATGGCGACCACTATCATTATTACAATGGTAAGGTTCCTTATGACGCTATCATCAG 240
          |||||||||||||||||||||||||||||||||||||||||||||||| || ||  ||||
Db    7628 TTCACATGGCGACCACTATCATTATTACAATGGTAAGGTTCCTTATGATGCGATTTTCAG 7569

Qy     241 TGAAGAATTACTCATGAAAGATCCAAACTATAAGCTAAAAGATGAGGATATTGTTAATGA 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    7568 TGAAGAATTACTCATGAAAGATCCAAACTATAAGCTAAAAGATGAGGATATTGTTAATGA 7509
```

```
Qy     301 GGTCAAGGGTGGATATGTTATCAAGGTAGATGGAAAATACTATGTTTACCTTAAGGATGC 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    7508 GGTCAAGGGTGGATATGTTATCAAGGTAGATGGAAAATACTATGTTTACCTTAAGGATGC 7449

Qy     361 TGCCCACGCGGATAACGTCCGTACAAAAGAGGAAATCAATCGACAAAAACAAGAGCATAG 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||| |||||
Db    7448 TGCCCACGCGGATAACGTCCGTACAAAAGAGGAAATCAATCGACAAAAACAAGAACATAG 7389

Qy     421 TCAACATCGTGAAGGTGGAACTCCAAGAAACGATGGTGCTGTTGCCTTGGCACGTTCGCA 480
           ||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db    7388 TCAACATCGTGAAGGTGGGACTCCAAGAAACGATGGTGCTGTTGCCTTGGCACGTTCGCA 7329

Qy     481 AGGACGCTATACTACAGATGATGGTTATATCTTTAATGCTTCTGATATCATAGAGGATAC 540
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    7328 AGGACGCTATACTACAGATGATGGTTATATCTTTAATGCTTCTGATATCATAGAGGATAC 7269

Qy     541 TGGTGATGCTTATATCGTTCCTCATGGAGATCATTACCATTACATTCCTAAGAATGAGTT 600
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    7268 TGGTGATGCTTATATCGTTCCTCATGGAGATCATTACCATTACATTCCTAAGAATGAGTT 7209

Qy     601 ATCAGCTAGCGAGTTGGCTGCTGCAGAAGCCTTCCTATCTGGTCGAGGAAATCTGTCAAA 660
           ||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db    7208 ATCAGCTAGCGAGTTGGCTGCTGCAGAAGCCTTCCTATCTGGTCGGGGAAATCTGTCAAA 7149

Qy     661 TTCAAGAACCTATCGCCGACAAAATAGCGATAACACTTCAAGAACAAACTGGGTACCTTC 720
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    7148 TTCAAGAACCTATCGCCGACAAAATAGCGATAACACTTCAAGAACAAACTGGGTACCTTC 7089

Qy     721 TGTAAGCAATCCAGGAACTACAAATACTAACACAAGCAACAACAGCAACACTAACAGTCA 780
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    7088 TGTAAGCAATCCAGGAACTACAAATACTAACACAAGCAACAACAGCAACACTAACAGTCA 7029

Qy     781 AGCAAGTCAAAGTAATGACATTGATAGTCTCTTGAAACAGCTCTACAAACTGCCTTTGAG 840
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    7028 AGCAAGTCAAAGTAATGACATTGATAGTCTCTTGAAACAGCTCTACAAACTGCCTTTGAG 6969

Qy     841 TCAACGACATGTAGAATCTGATGGCCTTGTCTTTGATCCAGCACAAATCACAAGTCGAAC 900
           |||||||||||||||||||||||||||| || || ||||| ||||||||||||||||||||
Db    6968 TCAACGACATGTAGAATCTGATGGCCTTGTTTTCGACCCAGCGCAAATCACAAGTCGAAC 6909

Qy     901 AGCTAGAGGTGTTGCAGTGCCACACGGAGATCATTACCACTTCATCCCTTACTCTCAAAT 960
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    6908 AGCTAGAGGTGTTGCAGTGCCACACGGAGATCATTACCACTTCATCCCTTACTCTCAAAT 6849

Qy     961 GTCTGAATTGGAAGAACGAATCGCTCGTATTATTCCCCTTCGTTATCGTTCAAACCATTG 1020
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    6848 GTCTGAATTGGAAGAACGAATCGCTCGTATTATTCCCCTTCGTTATCGTTCAAACCATTG 6789

Qy    1021 GGTACCAGATTCAAGGCCAGAACAACCAAGTCCACAACCGACTCCGGAACCTAGTCCAGG 1080
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    6788 GGTACCAGATTCAAGGCCAGAACAACCAAGTCCACAACCGACTCCGGAACCTAGTCCAGG 6729

Qy    1081 CCCGCAACCTGCACCAAATCTTAAAATAGACTCAAATTCTTCTTTGGTTAGTCAGCTGGT 1140
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    6728 CCCGCAACCTGCACCAAATCTTAAAATAGACTCAAATTCTTCTTTGGTTAGTCAGCTGGT 6669

Qy    1141 ACGAAAAGTTGGGGAAGGATATGTATTCGAAGAAAAGGGCATCTCTCGTTATGTCTTTGC 1200
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    6668 ACGAAAAGTTGGGGAAGGATATGTATTCGAAGAAAAGGGCATCTCTCGTTATGTCTTTGC 6609

Qy    1201 GAAAGATTTACCATCTGAAACTGTTAAAAATCTTGAAAGCAAGTTATCAAAACAAGAGAG 1260
           |||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db    6608 GAAAGATTTACCATCTGAAACTGTTAAAAATCTTGAAAGTAAGTTATCAAAACAAGAGAG 6549

Qy    1261 TGTTTCACACACTTTAACTGCTAAAAAAGAAAATGTTGCTCCTCGTGACCAAGAATTTTA 1320
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    6548 TGTTTCACACACTTTAACTGCTAAAAAAGAAAATGTTGCTCCTCGTGACCAAGAATTTTA 6489

Qy    1321 TGATAAAGCATATAATCTGTTAACTGAGGCTCATAAAGCCTTGTTTGNAAATAAGGGTCG 1380
           |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db    6488 TGATAAAGCATATAATCTGTTAACTGAGGCTCATAAAGCCTTGTTTGAAAATAAGGGTCG 6429

Qy    1381 TAATTCTGATTTCCAAGCCTTAGACAAATTATTAGAACGCTTGAATGATGAATCGACTAA 1440
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    6428 TAATTCTGATTTCCAAGCCTTAGACAAATTATTAGAACGCTTGAATGATGAATCGACTAA 6369

Qy    1441 TAAAGAAAAATTGGTAGATGATTTATTGGCATTCCTAGCACCAATTACCCATCCAGAGCG 1500
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    6368 TAAAGAAAAATTGGTAGATGATTTATTGGCATTCCTAGCACCAATTACCCATCCAGAGCG 6309

Qy    1501 ACTTGGCAAACCAAATTCTCAAATTGAGTATACTGAAGACGAAGTTCGTATTGCTCAATT 1560
           ||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    6308 ACTTGGCAAGCCAAATTCTCAAATTGAGTATACTGAAGACGAAGTTCGTATTGCTCAATT 6249

Qy    1561 AGCTGATAAGTATACAACGTCAGATGGTTACATTTTTGATGAACATGATATAATCAGTGA 1620
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    6248 AGCTGATAAGTATACAACGTCAGATGGTTACATTTTTGATGAACATGATATAATCAGTGA 6189

Qy    1621 TGAAGGAGATGCATATGTAACGCCTCATATGGGCCATAGTCACTGGATTGGAAAAGATAG 1680
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||
Db    6188 TGAAGGAGATGCATATGTAACGCCTCATATGGGCCATAGTCACTGGATTGGAAAAGACAG 6129

Qy    1681 CCTTTCTGATAAGGAAAAAGTTGCAGCTCAAGCCTATACTAAAGAAAAAGGTATCCTACC 1740
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    6128 CCTTTCTGATAAGGAAAAAGTTGCAGCTCAAGCCTATACTAAAGAAAAAGGTATCCTACC 6069

Qy    1741 TCCATCTCCAGACGCAGATGTTAAAGCAAATCCAACTGGAGATAGTGCAGCAGCTATTTA 1800
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    6068 TCCATCTCCAGACGCAGATGTTAAAGCAAATCCAACTGGAGATAGTGCAGCAGCTATTTA 6009

Qy    1801 CAATCGTGTGAAAGGGGAAAAACGAATTCCACTCGTTCGACTTCCATATATGGTTGAGCA 1860
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    6008 CAATCGTGTGAAAGGGGAAAAACGAATTCCACTCGTTCGACTTCCATATATGGTTGAGCA 5949

Qy    1861 TACAGTTGAGGTTAAAAACGGTAATTTGATTATTCCTCATAAGGATCATTACCATAATAT 1920
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    5948 TACAGTTGAGGTTAAAAACGGTAATTTGATTATTCCTCATAAGGATCATTACCATAATAT 5889

Qy    1921 TAAATTTGCTTGGTTTGATGATCACACATACAAAGCTCCAAATGGCTATACCTTGGAAGA 1980
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    5888 TAAATTTGCTTGGTTTGATGATCACACATACAAAGCTCCAAATGGCTATACCTTGGAAGA 5829

Qy    1981 TTTGTTTGCGACGATTAAGTACTACGTAGAACACCCTGACGAACGTCCACATTCTAATGA 2040
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    5828 TTTGTTTGCGACGATTAAGTACTACGTAGAACACCCTGACGAACGTCCACATTCTAATGA 5769

Qy    2041 TGGATGGGGCAATGCCAGTGAGCATGTGTTAGGCAAGAAAGACCACAGTGAAGATCCAAA 2100
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    5768 TGGATGGGGCAATGCCAGTGAGCATGTGTTAGGCAAGAAAGACCACAGTGAAGATCCAAA 5709

Qy    2101 TAAGAACTTCAAAGCGGATGAAGAGCCAGTAGAGGAAACACCTGCTGAGCCAGAAGTCCC 2160
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    5708 TAAGAACTTCAAAGCGGATGAAGAGCCAGTAGAGGAAACACCTGCTGAGCCAGAAGTCCC 5649

Qy    2161 TCAAGTAGAGACTGAAAAAGTAGAAGCCCAACTCAAAGAAGCAGAAGTTTTGCTTGCGAA 2220
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    5648 TCAAGTAGAGACTGAAAAAGTAGAAGCCCAACTCAAAGAAGCAGAAGTTTTGCTTGCGAA 5589

Qy    2221 AGTAACGGATTCTAGTCTGAAAGCCAATGCAACAGAAACTCTAGCTGGTTTACGAAATAA 2280
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    5588 AGTAACGGATTCTAGTCTGAAAGCCAATGCAACAGAAACTCTAGCTGGTTTACGAAATAA 5529

Qy    2281 TTTGACTCTTCAAATTATGGATAACAATAGTATCATGGCAGAAGCAGAAAAATTACTTGC 2340
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    5528 TTTGACTCTTCAAATTATGGATAACAATAGTATCATGGCAGAAGCAGAAAAATTACTTGC 5469

Qy    2341 GTTGTTAAAAGGAAGTAATCCTTCATCTGTAAGTAAGGAAAAAATAAAC 2389
           ||||||||||||||||||||||||||||||||||||||||||||||||||
Db    5468 GTTGTTAAAAGGAAGTAATCCTTCATCTGTAAGTAAGGAAAAAATAAAC 5420
```

RESULT 6
SPNEU1915
LOCUS SPNEU1915 20035 bp DNA linear HTG 11-JUL-2001
DEFINITION Streptococcus pneumoniae clone G54, *** SEQUENCING IN PROGRESS ***, in ordered pieces.

```
ACCESSION   AL449937
VERSION     AL449937.1  GI:11545162
KEYWORDS    HTG; HTGS_PHASE2.
SOURCE      Streptococcus pneumoniae.
  ORGANISM  Streptococcus pneumoniae
            Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
            Streptococcus.
REFERENCE   1  (bases 1 to 20035)
  AUTHORS   Dopazo,J., Mendoza,A., Herrero,J., Caldara,F., Humbert,Y.,
            Friedli,L., Guerrier,M., Grand-Schenk,E., Gandin,C., de
            Francesco,M., Polissi,A., Buell,G., Feger,G., Garcia,E., Peitsch,M.
            and Garcia-Bustos,J.F.
  TITLE     Annotated draft genomic sequence from a Streptococcus pneumoniae
            type 19F clinical isolate
  JOURNAL   Microb. Drug Resist. 7 (2), 99-125 (2001)
  MEDLINE   21335329
   PUBMED   11442348
REFERENCE   2  (bases 1 to 20035)
  AUTHORS   Dopazo,J., Mendoza,A., Herrero,J., Caldara,F., Polissi,A.,
            Humbert,Y., Friedli,L., Guerrier,M., Grand-Schenk,E., Gandin,C., de
            Francesco,M., Buell,G., Feger,G., Garcia,E., Peitsch,M. and
            Garcia-Bustos,J.F.
  TITLE     Direct Submission
  JOURNAL   Submitted (31-OCT-2000) Research Department, Glaxo Wellcome, S.A.,
            Severo Ochoa 2, 28760 Tres Cantos, SPAIN
COMMENT     * NOTE: This is a 'working draft' sequence.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
FEATURES             Location/Qualifiers
     source          1. .20035
                     /organism="Streptococcus pneumoniae"
                     /serotype="19F"
                     /db_xref="taxon:1313"
                     /clone="G54"
BASE COUNT     6369 a   3635 c   4108 g   5909 t     14 others
ORIGIN
```

Query Match 97.5%; Score 2328.2; DB 2; Length 20035;
Best Local Similarity 98.5%; Pred. No. 0;
Matches 2353; Conservative 6; Mismatches 30; Indels 1; Gaps 1;

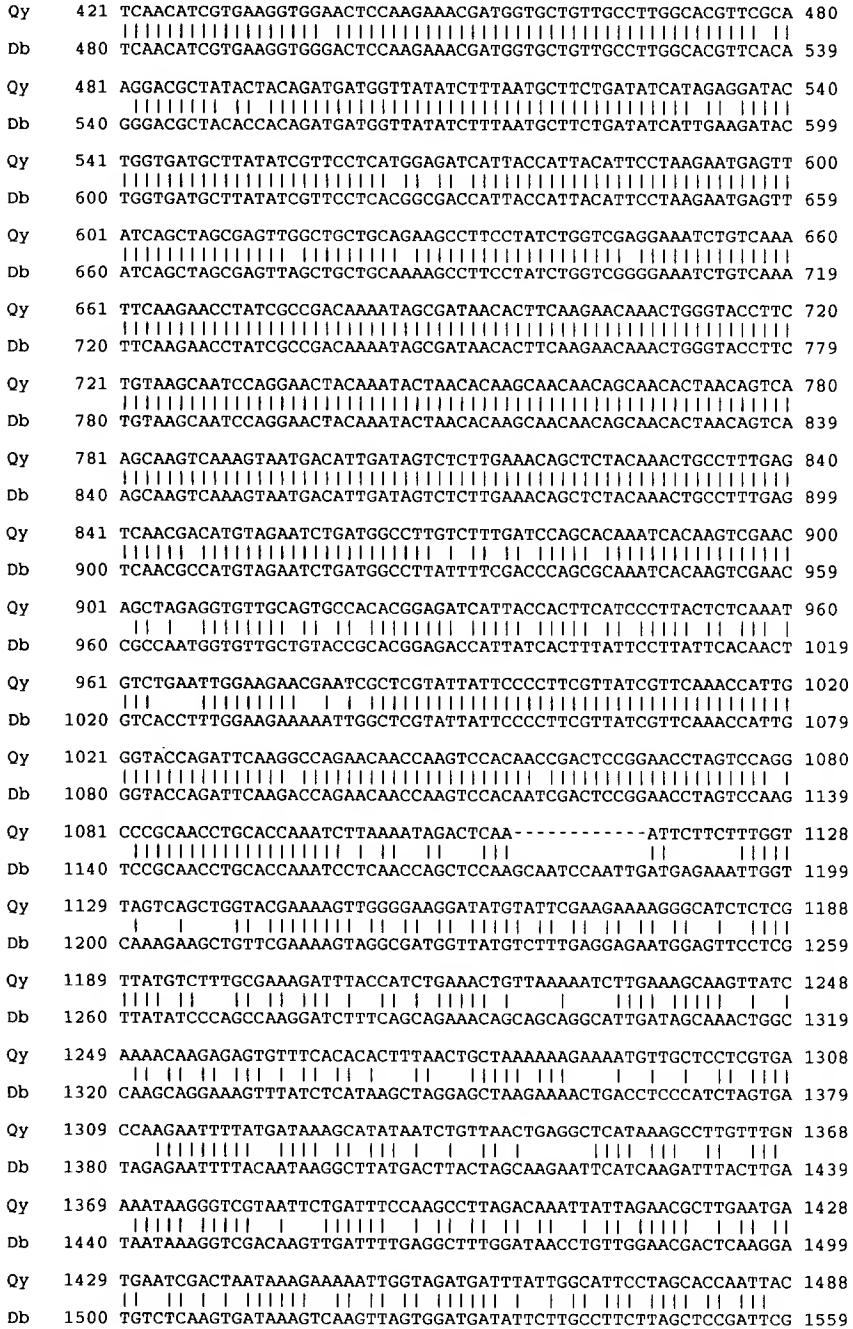
```
Qy       1 TTCTTACGAGTTGGGACTGTATCAAGCTAGAACGGTTAAGGAAAATAATCGTGTTTCCTA 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   17070 TTCTTACGAGTTGGGACTGTATCAAGCTAGAACGGTTAAGGAAAATAATCGTGTTTCCTA 17129

Qy      61 TATAGATGGAAAACAAGCGACGCAAAAAACGGAGAATTTGACTCCTGATGAGGTTAGCAA 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   17130 TATAGATGGAAAACAAGCGACGCAAAAAACGGAGAATTTGACTCCTGATGAGGTTAGCAA 17189

Qy     121 GCGTGAAGGAATCAATGCTGAGCAAATCGTCATCAAGATAACAGACCAAGGCTATGTCAC 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   17190 GCGTGAAGGAATCAATGCTGAGCAAATCGTCATCAAGATAACAGACCAAGGCTATGTCAC 17249

Qy     181 TTCACATGGCGACCACTATCATTATTACAATGGTAAGGTTCCTTATGACGCTATCATCAG 240
          |||||||||||||||||||||||||||||||||||||||||||||||||| || ||  ||||
Db   17250 TTCACATGGCGACCACTATCATTATTACAATGGTAAGGTTCCTTATGATGCGATTTTCAG 17309

Qy     241 TGAAGAATTACTCATGAAAGATCCAAACTATAAGCTAAAAGATGAGGATATTGTTAATGA 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   17310 TGAAGAATTACTCATGAAAGATCCAAACTATAAGCTAAAAGATGAGGATATTGTTAATGA 17369

Qy     301 GGTCAAGGGTGGATATGTTATCAAGGTAGATGGAAAATACTATGTTTACCTTAAGGATGC 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   17370 GGTCAAGGGTGGATATGTTATCAAGGTAGATGGAAAATACTATGTTTACCTTAAGGATGC 17429

Qy     361 TGCCCACGCGGATAACGTCCGTACAAAAGAGGAAATCAATCGACAAAAACAAGAGCATAG 420
          |||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||
Db   17430 TGCCCACGCGGATAACGTCCGTACAAAAGAGGAAATCAATCGACAAAAACAAGAACATAG 17489

Qy     421 TCAACATCGTGAAGGTGGAACTCCAAGAAACGATGGTGCTGTTGCCTTGGCACGTTCGCA 480
          |||||||||||||| |||| |||||||||||||||||||| || ||||| |||||||| ||
Db   17490 TCAACATCGTGAAAGTGGGACTCCAAGAAACGATGGTGCGGTAGCCTTTGCACGTTCACA 17549

Qy     481 AGGACGCTATACTACAGATGATGGTTATATCTTTAATGCTTCTGATATCATAGAGGATAC 540
          |||||||| || ||||||||||||||||||||||||||||||||||||||||| ||||| ||
Db   17550 GGGACGCTACACCACAGATGATGGTTATATCTTTAATGCTTCTGATATCATTGAGGACAC 17609

Qy     541 TGGTGATGCTTATATCGTTCCTCATGGAGATCATTACCATTACATTCCTAAGAATGAGTT 600
          |||||||||||||||||||||||||| || ||||||||||||||||||||||||||||||
Db   17610 GGGTGATGCTTATATCGTTCCTCATGGCGACCATTACCATTACATTCCTAAGAATGAGTT 17669

Qy     601 ATCAGCTAGCGAGTTGGCTGCTGCAGAAGCCTTCCTATCTGGTCGAGGAAATCTGTCAAA 660
          |||||||||||||||:|||||||||||||||||||||||||||||| ||||||||||||||
Db   17670 ATCAGCTAGCGAGTTRGCTGCTGCAGAAGCCTTCCTATCTGGTCGGGGAAATCTGTCAAA 17729

Qy     661 TTCAAGAACCTATCGCCGACAAAATAGCGATAACACTTCAAGAACAAACTGGGTACCTTC 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   17730 TTCAAGAACCTATCGCCGACAAAATAGCGATAACACTTCAAGAACAAACTGGGTACCTTC 17789

Qy     721 TGTAAGCAATCCAGGAACTACAAATACTAACACAAGCAACAACAGCAACACTAACAGTCA 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   17790 TGTAAGCAATCCAGGAACTACAAATACTAACACAAGCAACAACAGCAACACTAACAGTCA 17849

Qy     781 AGCAAGTCAAAGTAATGACATTGATAGTCTCTTGAAACAGCTCTACAAACTGCCTTTGAG 840
          |:|||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   17850 ARCAAGTCAAAGTAATGACATTGATAGTCTCTTGAAACAGCTCTACAAACTGCCTTTGAG 17909

Qy     841 TCAACGACATGTAGAATCTGATGGCCTTGTCTTTGATCCAGCACAAATCACAAGTCGAAC 900
          ||||||:|||||||||||||||||||||| || || ||||| ||||||||||||||||||
Db   17910 TCAACGMCATGTAGAATCTGATGGCCTTGTTTTCGACCCAGCGCAAATCACAAGTCGAAC 17969

Qy     901 AGCTAGAGGTGTTGCAGTGCCACACGGAGATCATTACCACTTCATCCCTTACTCTCAAAT 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   17970 AGCTAGAGGTGTTGCAGTGCCACACGGAGATCATTACCACTTCATCCCTTACTCTCAAAT 18029

Qy     961 GTCTGAATTGGAAGAACGAATCGCTCGTATTATTCCCCTTCGTTATCGTTCAAACCATTG 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   18030 GTCTGAATTGGAAGAACGAATCGCTCGTATTATTCCCCTTCGTTATCGTTCAAACCATTG 18089

Qy    1021 GGTACCAGATTCAAGGCCAGAACAACCAAGTCCACAACCGACTCCGGAACCTAGTCCAGG 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   18090 GGTACCAGATTCAAGGCCAGAACAACCAAGTCCACAACCGACTCCGGAACCTAGTCCAGG 18149

Qy    1081 CCCGCAACCTGCACCAAATCTTAAAATAGACTCAAATTCTTCTTTGGTTAGTCAGCTGGT 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   18150 CCCGCAACCTGCACCAAATCTTAAAATAGACTCAAATTCTTCTTTGGTTAGTCAGCTGGT 18209

Qy    1141 ACGAAAAGTTGGGGAAGGATATGTATTCGAAGAAAAGGGCATCTCTCGTTATGTCTTTGC 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   18210 ACGAAAAGTTGGGGAAGGATATGTATTCGAAGAAAAGGGCATCTCTCGTTATGTCTTTGC 18269

Qy    1201 GAAAGATTTACCATCTGAAACTGTTAAAAATCTTGAAAGCAAGTTATCAAAACAAGAGAG 1260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   18270 GAAAGATTTACCATCTGAAACTGTTAAAAATCTTGAAAGCAAGTTATCAAAACAAGAGAG 18329

Qy    1261 TGTTTCACACACTTTAACTGCTAAAAAAGAAAATGTTGCTCCTCGTGACCAAGAATTTTA 1320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||:
Db   18330 TGTTTCACACACTTTAACTGCTAAAAAAGAAAATGTTGCTCCTCGTGACCAAGAATTTTW 18389

Qy    1321 TGATAAAGCATATAATCTGTTAACTGAGGCTCATAAAGCCTTGTTTGNAAATAAGGGTCG 1380
          |||||||||||||||||||||||||||||||||||||||||:||||||||| ||||||||||
Db   18390 TGATAAAGCATATAATCTGTTAACTGAGGCTCATAAASCCTTGTTTGAAAATAAGGGTCG 18449

Qy    1381 TAATTCTGATTTCCAAGCCTTAGACAAATTATTAGAACGCTTGAATGATGAATCGAC-TA 1439
          |||||||||||||||:|||| |||||||| ||||||||||||||||||||||||||| ||
Db   18450 TAATTCTGATTTCCAASCCTTGGACAAATTTTTAGAACGCTTGAATGATGAATCGACTTA 18509

Qy    1440 ATAAAGAAAAATTGGTAGATGATTTATTGGCATTCCTAGCACCAATTACCCATCCAGAGC 1499
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   18510 ATAAAGAAAAATTGGTAGATGATTTATTGGCATTCCTAGCACCAATTACCCATCCAGAGC 18569

Qy    1500 GACTTGGCAAACCAAATTCTCAAATTGAGTATACTGAAGACGAAGTTCGTATTGCTCAAT 1559
          |||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   18570 GACTTGGCAAGCCAAATTCTCAAATTGAGTATACTGAAGACGAAGTTCGTATTGCTCAAT 18629
```

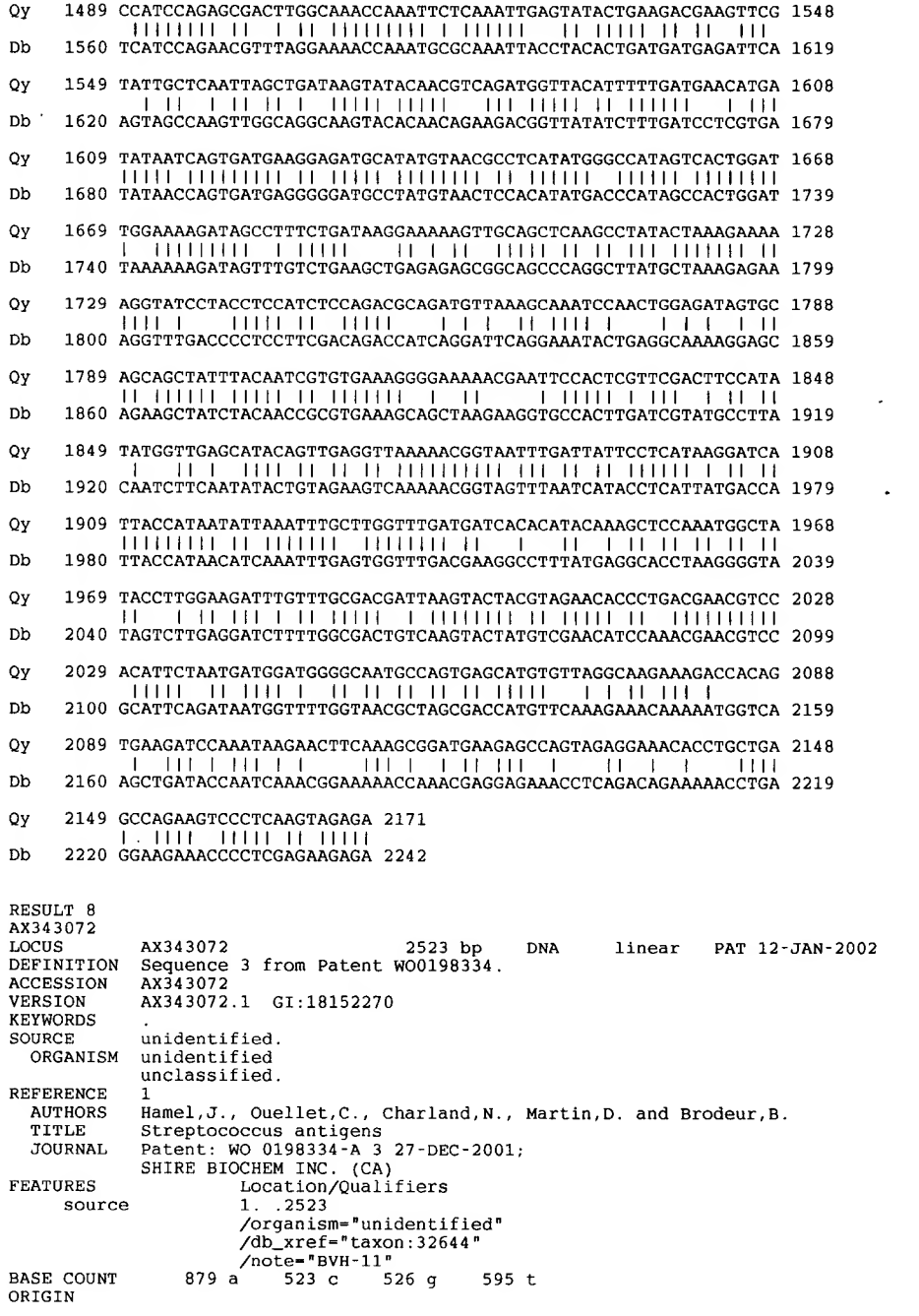
```
Qy   1560 TAGCTGATAAGTATACAACGTCAGATGGTTACATTTTTGATGAACATGATATAATCAGTG 1619
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  18630 TAGCTGATAAGTATACAACGTCAGATGGTTACATTTTTGATGAACATGATATAATCAGTG 18689

Qy   1620 ATGAAGGAGATGCATATGTAACGCCTCATATGGGCCATAGTCACTGGATTGGAAAAGATA 1679
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db  18690 ATGAAGGAGATGCATATGTAACGCCTCATATGGGCCATAGTCACTGGATTGGAAAAGACA 18749

Qy   1680 GCCTTTCTGATAAGGAAAAAGTTGCAGCTCAAGCCTATACTAAAGAAAAAGGTATCCTAC 1739
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  18750 GCCTTTCTGATAAGGAAAAAGTTGCAGCTCAAGCCTATACTAAAGAAAAAGGTATCCTAC 18809

Qy   1740 CTCCATCTCCAGACGCAGATGTTAAAGCAAATCCAACTGGAGATAGTGCAGCAGCTATTT 1799
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  18810 CTCCATCTCCAGACGCAGATGTTAAAGCAAATCCAACTGGAGATAGTGCAGCAGCTATTT 18869

Qy   1800 ACAATCGTGTGAAAGGGGAAAAACGAATTCCACTCGTTCGACTTCCATATATGGTTGAGC 1859
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  18870 ACAATCGTGTGAAAGGGGAAAAACGAATTCCACTCGTTCGACTTCCATATATGGTTGAGC 18929

Qy   1860 ATACAGTTGAGGTTAAAAACGGTAATTTGATTATTCCTCATAAGGATCATTACCATAATA 1919
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db  18930 ATACAGTTGAGGTTAAAAACGGTAATTTGATTATTCCTCATAAGGATCATTACCATAACA 18989

Qy   1920 TTAAATTTGCTTGGTTTGATGATCACACATACAAAGCTCCAAATGGCTATACCTTGGAAG 1979
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  18990 TTAAATTTGCTTGGTTTGATGATCACACATACAAAGCTCCAAATGGCTATACCTTGGAAG 19049

Qy   1980 ATTTGTTTGCGACGATTAAGTACTACGTAGAACACCCTGACGAACGTCCACATTCTAATG 2039
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  19050 ATTTGTTTGCGACGATTAAGTACTACGTAGAACACCCTGACGAACGTCCACATTCTAATG 19109

Qy   2040 ATGGATGGGGCAATGCCAGTGAGCATGTGTTAGGCAAGAAAGACCACAGTGAAGATCCAA 2099
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  19110 ATGGATGGGGCAATGCCAGTGAGCATGTGTTAGGCAAGAAAGACCACAGTGAAGATCCAA 19169

Qy   2100 ATAAGAACTTCAAAGCGGATGAAGAGCCAGTAGAGGAAACACCTGCTGAGCCAGAAGTCC 2159
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  19170 ATAAGAACTTCAAAGCGGATGAAGAGCCAGTAGAGGAAACACCTGCTGAGCCAGAAGTCC 19229

Qy   2160 CTCAAGTAGAGACTGAAAAAGTAGAAGCCCAACTCAAAGAAGCAGAAGTTTTGCTTGCGA 2219
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  19230 CTCAAGTAGAGACTGAAAAAGTAGAAGCCCAACTCAAAGAAGCAGAAGTTTTGCTTGCGA 19289

Qy   2220 AAGTAACGGATTCTAGTCTGAAAGCCAATGCAACAGAAACTCTAGCTGGTTTACGAAATA 2279
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  19290 AAGTAACGGATTCTAGTCTGAAAGCCAATGCAACAGAAACTCTAGCTGGTTTACGAAATA 19349

Qy   2280 ATTTGACTCTTCAAATTATGGATAACAATAGTATCATGGCAGAAGCAGAAAAATTACTTG 2339
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  19350 ATTTGACTCTTCAAATTATGGATAACAATAGTATCATGGCAGAAGCAGAAAAATTACTTG 19409

Qy   2340 CGTTGTTAAAAGGAAGTAATCCTTCATCTGTAAGTAAGGAAAAAATAAAC 2389
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Db  19410 CGTTGTTAAAAGGAAGTAATCCTTCATCTGTAAGTAAGGAAAAAATAAAC 19459
```

```
RESULT 7
AF340221
LOCUS       AF340221               2535 bp    DNA     linear   BCT 22-MAY-2001
DEFINITION  Streptococcus pneumoniae PhpA (phpA) gene, complete cds.
ACCESSION   AF340221
VERSION     AF340221.1  GI:13447093
KEYWORDS    .
SOURCE      Streptococcus pneumoniae.
  ORGANISM  Streptococcus pneumoniae
            Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
            Streptococcus.
REFERENCE   1  (bases 1 to 2535)
  AUTHORS   Zhang,Y., Masi,A.W., Barniak,V., Mountzouros,K., Hostetter,M.K. and
            Green,B.A.
  TITLE     Recombinant PhpA protein, a unique histidine motif-containing
            protein from Streptococcus pneumoniae, protects mice against
            intranasal pneumococcal challenge
  JOURNAL   Infect. Immun. 69 (6), 3827-3836 (2001)
  MEDLINE   21246685
   PUBMED   11349048
REFERENCE   2  (bases 1 to 2535)
  AUTHORS   Zhang,Y., Masi,A., Barniak,V., Mountzouros,K., Hostetter,M. and
            Green,B.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-JAN-2001) Department of Bacteriology, Wyeth Lederle
            Vaccines, 211 Bailey Road, West Henrietta, NY 14586, USA
FEATURES             Location/Qualifiers
     source          1. .2535
                     /organism="Streptococcus pneumoniae"
                     /db_xref="taxon:1313"
     gene            1. .2535
                     /gene="phpA"
     CDS             1. .2535
                     /gene="phpA"
                     /note="histidine motif-containing protein"
                     /codon_start=1
                     /transl_table=11
                     /product="PhpA"
                     /protein_id="AAK26629.1"
                     /db_xref="GI:13447094"
                     /translation="MKINKKYLVGSAAALILSVCSYELGLYQARTVKENNRVSYIDGK
                     QATQKTENLTPDEVSKREGINAEQIVIKITDQGYVTSHGDHYHYYNGKVPYDAIISEE
                     LLMKDPNYQLKDEDIISEIKGGYVIKVDGKYYVYLKDAAHADNVRTKEEINRQKQEHS
                     QHREGGTPRNDGAVALARSQGRYTTDDGYIFNASDIIEDTGDAYIVPHGDHYHYIPKN
                     ELSASELAAAKAFLSGRGNLSNSRTYRRQNSDNTSRTNWVPSVSNPGTTNTNTSNNSN
                     TNSQASQSNDIDSLLKQLYKLPLSQRHVESDGLIFDPAQITSRTANGVAVPHGDHYHF
                     IPYSQLSPLEEKLARIIPLRYRSNHWVPDSRPEQPSPQSTPEPSPSPQPAPNPQPAPS
                     NPIDEKLVKEAVRKVGDGYVFEENGVPRYIPAKDLSAETAAGIDSKLAKQESLSHKLG
                     AKKTDLPSSDREFYNKAYDLLARIHQDLLDNKGRQVDFEALDNLLERLKDVSSDKVKL
                     VDDILAFLAPIRHPERLGKPNAQITYTDDEIQVAKLAGKYTTEDGYIFDPRDITSDEG
                     DAYVTPHMTHSHWIKKDSLSEAERAAAQAYAKEKGLTPPSTDHQDSGNTEAKGAEAIY
                     NRVKAAKKVPLDRMPYNLQYTVEVKNGSLIIPHYDHYHNIKFEWFDEGLYEAPKGYSL
                     EDLLATVKYYVEHPNERPHSDNGFGNASDHVQRNKNGQADTNQTEKPNEEKPQTEKPE
                     EETPREEKPQSEKPESPKPTEEPEEESPEESPEESEEPQVETEKVKEKLREAEDLLGK
                     IQNPIIKSNAKETLTGLKNNLLFGTQDNNTIMAEAEKLLALLKESK"
BASE COUNT      880 a    518 c    533 g    604 t
ORIGIN
```

Query Match 60.0%; Score 1434.4; DB 1; Length 2535;
Best Local Similarity 79.0%; Pred. No. 3.3e-296;
Matches 1724; Conservative 0; Mismatches 447; Indels 12; Gaps 1;

```
Qy      1 TTCTTACGAGTTGGGACTGTATCAAGCTAGAACGGTTAAGGAAAATAATCGTGTTTCCTA 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     60 TTCTTACGAGTTGGGACTGTATCAAGCTAGAACGGTTAAGGAAAATAATCGTGTTTCCTA 119

Qy     61 TATAGATGGAAAACAAGCGACGCAAAAAACGGAGAATTTGACTCCTGATGAGGTTAGCAA 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    120 TATAGATGGAAAACAAGCGACGCAAAAAACGGAGAATTTGACTCCTGATGAGGTTAGCAA 179

Qy    121 GCGTGAAGGAATCAATGCTGAGCAAATCGTCATCAAGATAACAGACCAAGGCTATGTCAC 180
          |||||||||||||||||  || ||||||||||||| ||||||||||||||||||||||||
Db    180 GCGTGAAGGAATCAATGCCGAACAAATCGTCATCAAAATAACAGACCAAGGCTATGTCAC 239

Qy    181 TTCACATGGCGACCACTATCATTATTACAATGGTAAGGTTCCTTATGACGCTATCATCAG 240
          |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    240 TTCACATGGGGACCACTATCATTATTACAATGGTAAGGTTCCTTATGACGCTATCATCAG 299

Qy    241 TGAAGAATTACTCATGAAAGATCCAAACTATAAGCTAAAAGATGAGGATATTGTTAATGA 300
          ||||||||| || ||||| |||||||||||  ||||||| || ||||||||| | | |||
Db    300 TGAAGAATTGCTGATGAAGGATCCAAACTACCAGCTAAAGGACGAGGATATTATCAGTGA 359

Qy    301 GGTCAAGGGTGGATATGTTATCAAGGTAGATGGAAAATACTATGTTTACCTTAAGGATGC 360
            |||||||||| |||| || ||||||||||||||||||||||||||||||||||||||||
Db    360 AATCAAGGGTGGTTATGTGATTAAGGTAGATGGAAAATACTATGTTTACCTTAAGGATGC 419

Qy    361 TGCCCACGCGGATAACGTCCGTACAAAAGAGGAAATCAATCGACAAAAACAAGAGCATAG 420
          || || |||||||| |||||||||||||| ||||||||||||||||||||||||||||||
Db    420 AGCTCATGCGGATAATGTCCGTACAAAAGAAGAAATCAATCGACAAAAACAAGAGCATAG 479
```

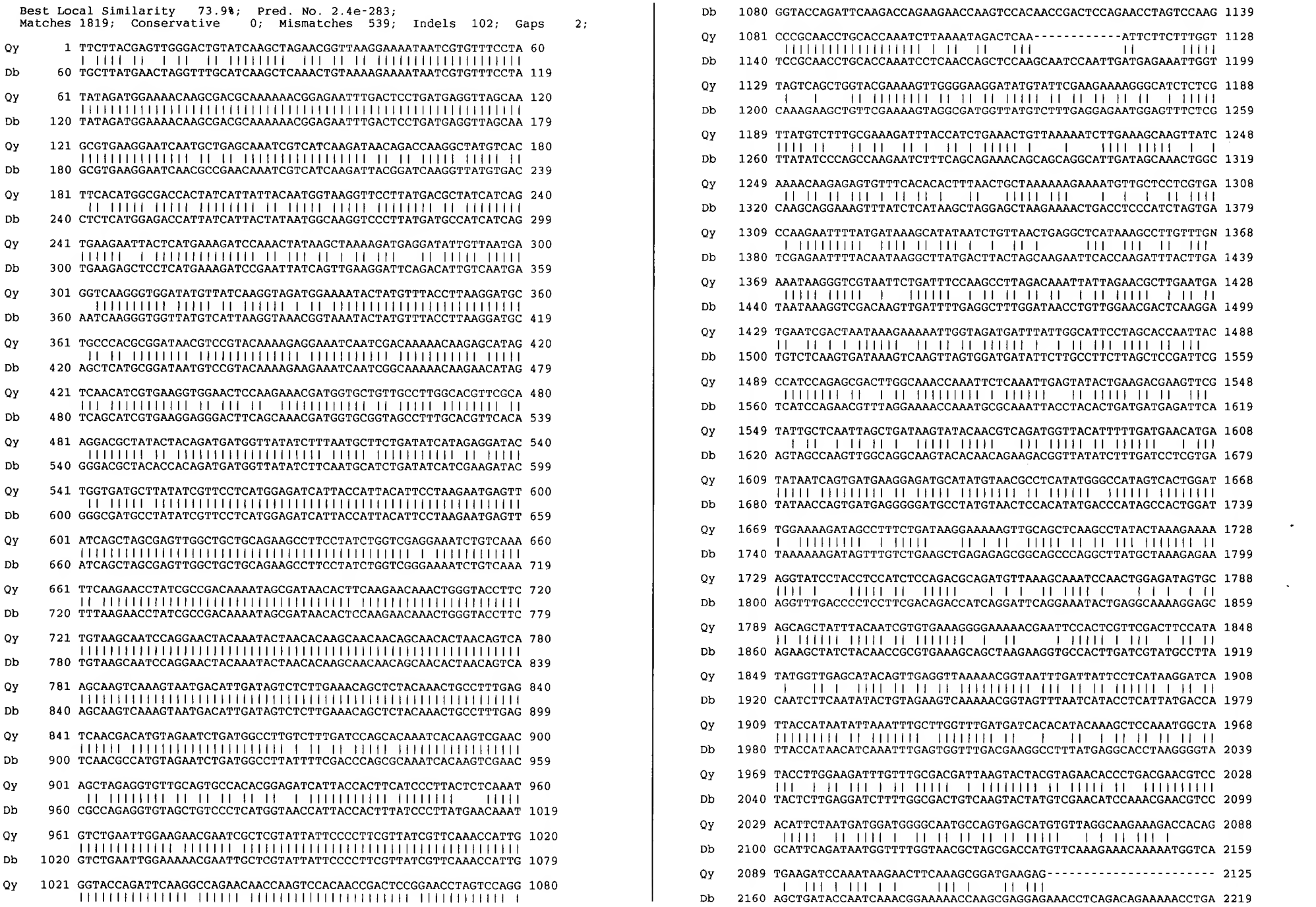




RESULT 8
AX343072
LOCUS AX343072 2523 bp DNA linear PAT 12-JAN-2002
DEFINITION Sequence 3 from Patent WO0198334.
ACCESSION AX343072
VERSION AX343072.1 GI:18152270
KEYWORDS .
SOURCE unidentified.
ORGANISM unidentified
unclassified.
REFERENCE 1
AUTHORS Hamel,J., Ouellet,C., Charland,N., Martin,D. and Brodeur,B.
TITLE Streptococcus antigens
JOURNAL Patent: WO 0198334-A 3 27-DEC-2001;
SHIRE BIOCHEM INC. (CA)
FEATURES Location/Qualifiers
source 1. .2523
/organism="unidentified"
/db_xref="taxon:32644"
/note="BVH-11"
BASE COUNT 879 a 523 c 526 g 595 t
ORIGIN

Query Match 57.5%; Score 1374.2; DB 6; Length 2523;

Best Local Similarity 73.9%; Pred. No. 2.4e-283;
Matches 1819; Conservative 0; Mismatches 539; Indels 102; Gaps 2;



```
Qy   2126 ------------------------------------------------------------ 2125

Db   2220 GGAAGAAACCCCTCGAGAAGAGAAACCACAAAGCGAGAAACCAGAGTCTCCAAAACCAAC 2279

Qy   2126 -------CCAGTAGAGGAAACACCTGCTGAGCCAGAAGTCCCTCAAGTAGAGACTGAAAA 2178
                 |||| ||| ||| |||| |  ||  |||||  ||||| || ||||||||||
Db   2280 AGAGGAACCAGAAGAAGAATCACCAGAGGAATCAGAAGAACCTCAGGTCGAGACTGAAAA 2339

Qy   2179 AGTAGAAGCCCAACTCAAAGAAGCAGAAGTTTTGCTTGCGAAAGTAACGGATTCTAGTCT 2238
           || ||||   |||| | ||| || |||| ||| ||||  ||| |   |||| | | | |
Db   2340 GGTTGAAGAAAAACTGAGAGAGGCTGAAGATTTACTTGGAAAAATCCAGGATCCAATTAT 2399

Qy   2239 GAAAGCCAATGCAACAGAAACTCTAGCTGGTTTACGAAATAATTTGACTCTTCAAATTAT 2298
           ||  |||||||| | ||| |||||  | || |||  ||||||||     ||   |
Db   2400 CAAGTCCAATGCCAAAGAGACTCTCACAGGATTAAAAAATAATTTACTATTTGGCACCCA 2459

Qy   2299 GGATAACAATAGTATCATGGCAGAAGCAGAAAAATTACTTGCGTTGTTAAAAGGAAGTAA 2358
          ||| ||||||| ||| |||||||||| |||||| || | || || ||||| |  |||||
Db   2460 GGACAACAATACTATTATGGCAGAAGCTGAAAAACTATTGGCTTTATTAAAGGAGAGTAA 2519


RESULT 9
AX343073
LOCUS       AX343073               2647 bp    DNA     linear   PAT 12-JAN-2002
DEFINITION  Sequence 4 from Patent WO0198334.
ACCESSION   AX343073
VERSION     AX343073.1  GI:18152271
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1
  AUTHORS   Hamel,J., Ouellet,C., Charland,N., Martin,D. and Brodeur,B.
  TITLE     Streptococcus antigens
  JOURNAL   Patent: WO 0198334-A 4 27-DEC-2001;
            SHIRE BIOCHEM INC. (CA)
FEATURES             Location/Qualifiers
     source          1. .2647
                     /organism="unidentified"
                     /db_xref="taxon:32644"
                     /note="BVH-11"
BASE COUNT      934 a    538 c    556 g    619 t
ORIGIN

  Query Match             57.5%;  Score 1374.2;  DB 6;  Length 2647;
  Best Local Similarity   73.9%;  Pred. No. 2.4e-283;
  Matches 1819;  Conservative    0;  Mismatches  539;  Indels  102;  Gaps    2;

Qy      1 TTCTTACGAGTTGGGACTGTATCAAGCTAGAACGGTTAAGGAAAATAATCGTGTTTCCTA 60
          | |||| ||  | ||  || |||||||  ||| || || ||||||||||||||||||||
Db    104 TGCTTATGAACTAGGTTTGCATCAAGCTCAAACTGTAAAAGAAAATAATCGTGTTTCCTA 163

Qy     61 TATAGATGGAAAACAAGCGACGCAAAAAACGGAGAATTTGACTCCTGATGAGGTTAGCAA 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    164 TATAGATGGAAAACAAGCGACGCAAAAAACGGAGAATTTGACTCCTGATGAGGTTAGCAA 223

Qy    121 GCGTGAAGGAATCAATGCTGAGCAAATCGTCATCAAGATAACAGACCAAGGCTATGTCAC 180
          |||||||||||||||| || || |||||||||||||||| || || ||||| ||||| ||
Db    224 GCGTGAAGGAATCAACGCCGAACAAATCGTCATCAAGATTACGGATCAAGGTTATGTGAC 283

Qy    181 TTCACATGGCGACCACTATCATTATTACAATGGTAAGGTTCCTTATGACGCTATCATCAG 240
           || ||||| ||||| |||||||| || ||||| ||||| |||||||| || ||||||||
Db    284 CTCTCATGGAGACCATTATCATTACTATAATGGCAAGGTCCCTTATGATGCCATCATCAG 343

Qy    241 TGAAGAATTACTCATGAAAGATCCAAACTATAAGCTAAAAGATGAGGATATTGTTAATGA 300
          ||||||  | ||||||||||||||| || ||| || | || |||   || ||||| |||||
Db    344 TGAAGAGCTCCTCATGAAAGATCCGAATTATCAGTTGAAGGATTCAGACATTGTCAATGA 403

Qy    301 GGTCAAGGGTGGATATGTTATCAAGGTAGATGGAAAATACTATGTTTACCTTAAGGATGC 360
            |||||||||| ||||| || |||||| | || ||||||||||||||||||||||||||
Db    404 AATCAAGGGTGGTTATGTCATTAAGGTAAACGGTAAATACTATGTTTACCTTAAGGATGC 463

Qy    361 TGCCCACGCGGATAACGTCCGTACAAAAGAGGAAATCAATCGACAAAAACAAGAGCATAG 420
          || || |||||||| |||||||||||||| |||||||||| |||||||||| |||||
Db    464 AGCTCATGCGGATAATGTCCGTACAAAAGAAGAAATCAATCGGCAAAAACAAGAACATAG 523

Qy    421 TCAACATCGTGAAGGTGGAACTCCAAGAAACGATGGTGCTGTTGCCTTGGCACGTTCGCA 480
          ||| |||||||||| || ||| ||  |||||||||||| || ||||| |||||||| ||
Db    524 TCAGCATCGTGAAGGAGGGACTTCAGCAAACGATGGTGCGGTAGCCTTTGCACGTTCACA 583

Qy    481 AGGACGCTATACTACAGATGATGGTTATATCTTTAATGCTTCTGATATCATAGAGGATAC 540
          |||||||| || |||||||||||||||||||| ||||| |||||||||| || |||||
Db    584 GGGACGCTACACCACAGATGATGGTTATATCTTCAATGCATCTGATATCATCGAAGATAC 643

Qy    541 TGGTGATGCTTATATCGTTCCTCATGGAGATCATTACCATTACATTCCTAAGAATGAGTT 600
          || ||||| |||||||||||||||||||||||||||||||||||||||||||||||||||
Db    644 GGGCGATGCCTATATCGTTCCTCATGGAGATCATTACCATTACATTCCTAAGAATGAGTT 703

Qy    601 ATCAGCTAGCGAGTTGGCTGCTGCAGAAGCCTTCCTATCTGGTCGAGGAAATCTGTCAAA 660
          |||||||||||||||||||||||||||||||||||||||||||| | ||||||||||||
Db    704 ATCAGCTAGCGAGTTGGCTGCTGCAGAAGCCTTCCTATCTGGTCGGGAAAATCTGTCAAA 763

Qy    661 TTCAAGAACCTATCGCCGACAAAATAGCGATAACACTTCAAGAACAAACTGGGTACCTTC 720
          || |||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db    764 TTTAAGAACCTATCGCCGACAAAATAGCGATAACACTCCAAGAACAAACTGGGTACCTTC 823

Qy    721 TGTAAGCAATCCAGGAACTACAAATACTAACACAAGCAACAACAGCAACACTAACAGTCA 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    824 TGTAAGCAATCCAGGAACTACAAATACTAACACAAGCAACAACAGCAACACTAACAGTCA 883

Qy    781 AGCAAGTCAAAGTAATGACATTGATAGTCTCTTGAAACAGCTCTACAAACTGCCTTTGAG 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    884 AGCAAGTCAAAGTAATGACATTGATAGTCTCTTGAAACAGCTCTACAAACTGCCTTTGAG 943

Qy    841 TCAACGACATGTAGAATCTGATGGCCTTGTCTTTGATCCAGCACAAATCACAAGTCGAAC 900
          |||||| |||||||||||||||||||| | || || ||||| ||||||||||||||||||
Db    944 TCAACGCCATGTAGAATCTGATGGCCTTATTTTCGACCCAGCGCAAATCACAAGTCGAAC 1003

Qy    901 AGCTAGAGGTGTTGCAGTGCCACACGGAGATCATTACCACTTCATCCCTTACTCTCAAAT 960
          || |||||||| || || || || ||  | |||||||||| |||||||    |||||
Db   1004 CGCCAGAGGTGTAGCTGTCCCTCATGGTAACCATTACCACTTTATCCCTTATGAACAAAT 1063

Qy    961 GTCTGAATTGGAAGAACGAATCGCTCGTATTATTCCCCTTCGTTATCGTTCAAACCATTG 1020
          |||||||||||| ||||||| ||||||||||||||||||||||||||||||||||||||
Db   1064 GTCTGAATTGGAAAAACGAATTGCTCGTATTATTCCCCTTCGTTATCGTTCAAACCATTG 1123

Qy   1021 GGTACCAGATTCAAGGCCAGAACAACCAAGTCCACAACCGACTCCGGAACCTAGTCCAGG 1080
          |||||||||||||||| |||||| ||||||||||||||||||||| |||||||||||| |
Db   1124 GGTACCAGATTCAAGACCAGAAGAACCAAGTCCACAACCGACTCCAGAACCTAGTCCAAG 1183

Qy   1081 CCCGCAACCTGCACCAAATCTTAAAATAGACTCAA----------ATTCTTCTTTGGT 1128
          |||||||||||||||||| | ||  ||  |||            ||      |||||
Db   1184 TCCGCAACCTGCACCAAATCCTCAACCAGCTCCAAGCAATCCAATTGATGAGAAATTGGT 1243

Qy   1129 TAGTCAGCTGGTACGAAAAGTTGGGGAAGGATATGTATTCGAAGAAAAGGGCATCTCTCG 1188
            |   |    || |||||||| || || || ||||| || || || || ||  | |||||
Db   1244 CAAAGAAGCTGTTCGAAAAGTAGGCGATGGTTATGTCTTTGAGGAGAATGGAGTTTCTCG 1303

Qy   1189 TTATGTCTTTGCGAAAGATTTACCATCTGAAACTGTTAAAAATCTTGAAAGCAAGTTATC 1248
          |||| ||   || ||  || |  || | ||||| |    |    |||| |||||  | |
Db   1304 TTATATCCCAGCCAAGAATCTTTCAGCAGAAACAGCAGCAGGCATTGATAGCAAACTGGC 1363

Qy   1249 AAAACAAGAGAGTGTTTCACACACTTTAACTGCTAAAAAAGAAAATGTTGCTCCTCGTGA 1308
          || || || ||| | || ||   ||  |||| |||    |  |  |  || ||||
Db   1364 CAAGCAGGAAAGTTTATCTCATAAGCTAGGAGCTAAGAAAACTGACCTCCCATCTAGTGA 1423

Qy   1309 CCAAGAATTTTATGATAAAGCATATAATCTGTTAACTGAGGCTCATAAAGCCTTGTTTGN 1368
          | |||||||||  |||| || ||| |  |  || |      |||  |||  ||  |||
Db   1424 TCGAGAATTTTACAATAAGGCTTATGACTTACTAGCAAGAATTCACCAAGATTTACTTGA 1483

Qy   1369 AAATAAGGGTCGTAATTCTGATTTCCAAGCCTTAGACAAATTATTAGAACGCTTGAATGA 1428
          ||||| |||||  |  ||||||  | || || || ||  | || |||||  | || ||
Db   1484 TAATAAAGGTCGACAAGTTGATTTTGAGGCTTTGGATAACCTGTTGGAACGACTCAAGGA 1543

Qy   1429 TGAATCGACTAATAAAGAAAAATTGGTAGATGATTTATTGGCATTCCTAGCACCAATTAC 1488
```

```
Db    1544 TGTCTCAAGTGATAAAGTCAAGTTAGTGGATGATATTCTTGCCTTCTTAGCTCCGATTCG 1603

Qy    1489 CCATCCAGAGCGACTTGGCAAACCAAATTCTCAAATTGAGTATACTGAAGACGAAGTTCG 1548
Db    1604 TCATCCAGAACGTTTAGGAAAACCAAATGCGCAAATTACCTACACTGATGATGAGATTCA 1663

Qy    1549 TATTGCTCAATTAGCTGATAAGTATACAACGTCAGATGGTTACATTTTTGATGAACATGA 1608
Db    1664 AGTAGCCAAGTTGGCAGGCAAGTACACAACAGAAGACGGTTATATCTTTGATCCTCGTGA 1723

Qy    1609 TATAATCAGTGATGAAGGAGATGCATATGTAACGCCTCATATGGGCCATAGTCACTGGAT 1668
Db    1724 TATAACCAGTGATGAGGGGGATGCCTATGTAACTCCACATATGACCCATAGCCACTGGAT 1783

Qy    1669 TGGAAAAGATAGCCTTTCTGATAAGGAAAAAGTTGCAGCTCAAGCCTATACTAAAGAAAA 1728
Db    1784 TAAAAAAGATAGTTTGTCTGAAGCTGAGAGAGCGGCAGCCCAGGCTTATGCTAAAGAGAA 1843

Qy    1729 AGGTATCCTACCTCCATCTCCAGACGCAGATGTTAAAGCAAATCCAACTGGAGATAGTGC 1788
Db    1844 AGGTTTGACCCCTCCTTCGACAGACCATCAGGATTCAGGAAATACTGAGGCAAAAGGAGC 1903

Qy    1789 AGCAGCTATTTACAATCGTGTGAAAGGGGAAAAACGAATTCCACTCGTTCGACTTCCATA 1848
Db    1904 AGAAGCTATCTACAACCGCGTGAAAGCAGCTAAGAAGGTGCCACTTGATCGTATGCCTTA 1963

Qy    1849 TATGGTTGAGCATACAGTTGAGGTTAAAAACGGTAATTTGATTATTCCTCATAAGGATCA 1908
Db    1964 CAATCTTCAATATACTGTAGAAGTCAAAAACGGTAGTTTAATCATACCTCATTATGACCA 2023

Qy    1909 TTACCATAATATTAAATTTGCTTGGTTTGATGATCACACATACAAAGCTCCAAATGGCTA 1968
Db    2024 TTACCATAACATCAAATTTGAGTGGTTTGACGAAGGCCTTTATGAGGCACCTAAGGGGTA 2083

Qy    1969 TACCTTGGAAGATTTGTTTGCGACGATTAAGTACTACGTAGAACACCCTGACGAACGTCC 2028
Db    2084 TACTCTTGAGGATCTTTTGGCGACTGTCAAGTACTATGTCGAACATCCAAACGAACGTCC 2143

Qy    2029 ACATTCTAATGATGGATGGGGCAATGCCAGTGAGCATGTGTTAGGCAAGAAAGACCACAG 2088
Db    2144 GCATTCAGATAATGGTTTTGGTAACGCTAGCGACCATGTTCAAAGAAACAAAAATGGTCA 2203

Qy    2089 TGAAGATCCAAATAAGAACTTCAAAGCGGATGAAGAG---------------------- 2125
Db    2204 AGCTGATACCAATCAAACGGAAAAACCAAGCGAGGAGAAACCTCAGACAGAAAAACCTGA 2263

Qy    2126 ---------------------------------------------------------- 2125
Db    2264 GGAAGAAACCCCTCGAGAAGAGAAACCACAAAGCGAGAAACCAGAGTCTCCAAAACCAAC 2323

Qy    2126 -------CCAGTAGAGGAAACACCTGCTGAGCCAGAAGTCCCTCAAGTAGAGACTGAAAA 2178
Db    2324 AGAGGAACCAGAAGAAGAATCACCAGAGGAATCAGAAGAACCTCAGGTCGAGACTGAAAA 2383

Qy    2179 AGTAGAAGCCCAACTCAAAGAAGCAGAAGTTTTGCTTGCGAAAGTAACGGATTCTAGTCT 2238
Db    2384 GGTTGAAGAAAAACTGAGAGAGGCTGAAGATTTACTTGGAAAAATCCAGGATCCAATTAT 2443

Qy    2239 GAAAGCCAATGCAACAGAAACTCTAGCTGGTTTACGAAATAATTTGACTCTTCAAATTAT 2298
Db    2444 CAAGTCCAATGCCAAAGAGACTCTCACAGGATTAAAAAATAATTTACTATTTGGCACCCA 2503

Qy    2299 GGATAACAATAGTATCATGGCAGAAGCAGAAAAATTACTTGCGTTGTTAAAAGGAAGTAA 2358
Db    2504 GGACAACAATACTATTATGGCAGAAGCTGAAAAACTATTGGCTTTATTAAAGGAGAGTAA 2563
```

RESULT 10
AF318954
LOCUS AF318954 2457 bp DNA linear BCT 11-FEB-2001
DEFINITION Streptococcus pneumoniae pneumococcal histidine triad protein B precursor (phtB) gene, partial cds.

ACCESSION AF318954
VERSION AF318954.1 GI:12744741
KEYWORDS .
SOURCE Streptococcus pneumoniae.
ORGANISM Streptococcus pneumoniae
Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Streptococcus.
REFERENCE 1 (bases 1 to 2457)
AUTHORS Adamou,J.E., Heinrichs,J.H., Erwin,A.L., Walsh,W., Gayle,T., Dormitzer,M., Dagan,R., Brewah,Y.A., Barren,P., Lathigra,R., Langermann,S., Koenig,S. and Johnson,S.
TITLE Identification and characterization of a novel family of pneumococcal proteins that are protective against sepsis
JOURNAL Infect. Immun. 69 (2), 949-958 (2001)
MEDLINE 21101045
PUBMED 11159990
REFERENCE 2 (bases 1 to 2457)
AUTHORS Adamou,J.E., Heinrichs,J.H., Erwin,A.L., Walsh,W., Dormitzer,M. and Johnson,S.
TITLE Direct Submission
JOURNAL Submitted (03-NOV-2000) Molecular Microbiology, MedImmune, Inc., 35 West Watkins Mill Road, Gaithersburg, MD 20878, USA
FEATURES Location/Qualifiers

```
     source          1. .2457
                     /organism="Streptococcus pneumoniae"
                     /serotype="4"
                     /db_xref="taxon:1313"
     gene            1. .>2457
                     /gene="phtB"
     CDS             1. .>2457
                     /gene="phtB"
                     /note="PhtB; protective pneumococcal antigen"
                     /codon_start=1
                     /transl_table=11
                     /product="pneumococcal histidine triad protein B
                     precursor"
                     /protein_id="AAK06759.1"
                     /db_xref="GI:12744742"
                     /translation="MKINKKYLAGSVAVLALSVCSYELGRYQAGQDKKESNRVAYIDG
                     DQAGQKAENLTPDEVSKREGINAEQIVIKITDQGYVTSHGDHYHYYNGKVPYDAIISE
                     ELLMKDPNYQLKDSDIVNEIKGGYVIKVNGKYYVYLKDAAHADNIRTKEEIKRQKQER
                     SHNHNSRADNAVAAARAQGRYTTDDGYIFNASDIIEDTGDAYIVPHGDHYHYIPKNEL
                     SASELAAAEAYWNGKQGSRPSSSSSYNANPAQPRLSENHNLTVTPTYHQNQGENISSL
                     LRELYAKPLSERHVESDGLIFDPAQITSRTARGVAVPHGNHYHFIPYEQMSELEKRIA
                     RIIPLRYRSNHWVPDSRPEEPSPQPTPEPSPSPQPAPSNPIDGKLVKEAVRKVGDGYV
                     FEENGVSRYIPAKDLSAETAAGIDSKLAKQESLSHKLGTKKTDLPSSDREFYNKAYDL
                     LARIHQDLLDNKGRQVDFEALDNLLERLKDVSSDKVKLVEDILAFLAPIRHPERLGKP
                     NAQITYTDDEIQVAKLAGKYTAEDGYIFDPRDITSDEGDAYVTPHMTHSHWIKKDSLS
                     EAERAAAQAYAXEKGLTPPSTDHQDSGNTEAKGAEAIYNRVKAAKKVPLDRMPYNLQY
                     TVEVKNGSLIIPHYDHYHNIKFEWFDEGLYEAPKGYTLEDLLATVKYYVEHPNERPHS
                     DNGFGNASDHVQRNKNGQADTNQTEKPSEEKPQTEKPEEETPREEKPQSEKPESPKPT
                     EEPEESPEESEEPQVETEKVEEKLREAEDLLGKIQDPIIKSNAKETLTGLKNNLLFGT
                     QDNNTIMAEAEKLLALLKESK"
     sig_peptide     1. .87
                     /gene="phtB"
     misc_feature    247. .264
                     /gene="phtB"
                     /note="Region: histidine triad"
     misc_feature    616. .633
                     /gene="phtB"
                     /note="Region: histidine triad"
     misc_feature    718. .792
                     /gene="phtB"
                     /note="Region: coiled-coil domain"
     misc_feature    940. .957
                     /gene="phtB"
                     /note="Region: histidine triad"
     misc_feature    1396. .1488
                     /gene="phtB"
                     /note="Region: coiled-coil domain"
     misc_feature    1657. .1674
                     /gene="phtB"
                     /note="Region: histidine triad"
```

```
     misc_feature    1909. .1926
                     /gene="phtB"
                     /note="Region: histidine triad"
     repeat_region   525. .714
                     /rpt_type=direct
     repeat_region   1578. .1758
                     /rpt_type=direct
BASE COUNT      832 a    496 c    533 g    593 t      3 others
ORIGIN
```

Query Match 41.5%; Score 991; DB 1; Length 2457;
Best Local Similarity 65.7%; Pred. No. 2e-201;
Matches 1609; Conservative 1; Mismatches 697; Indels 141; Gaps 6;

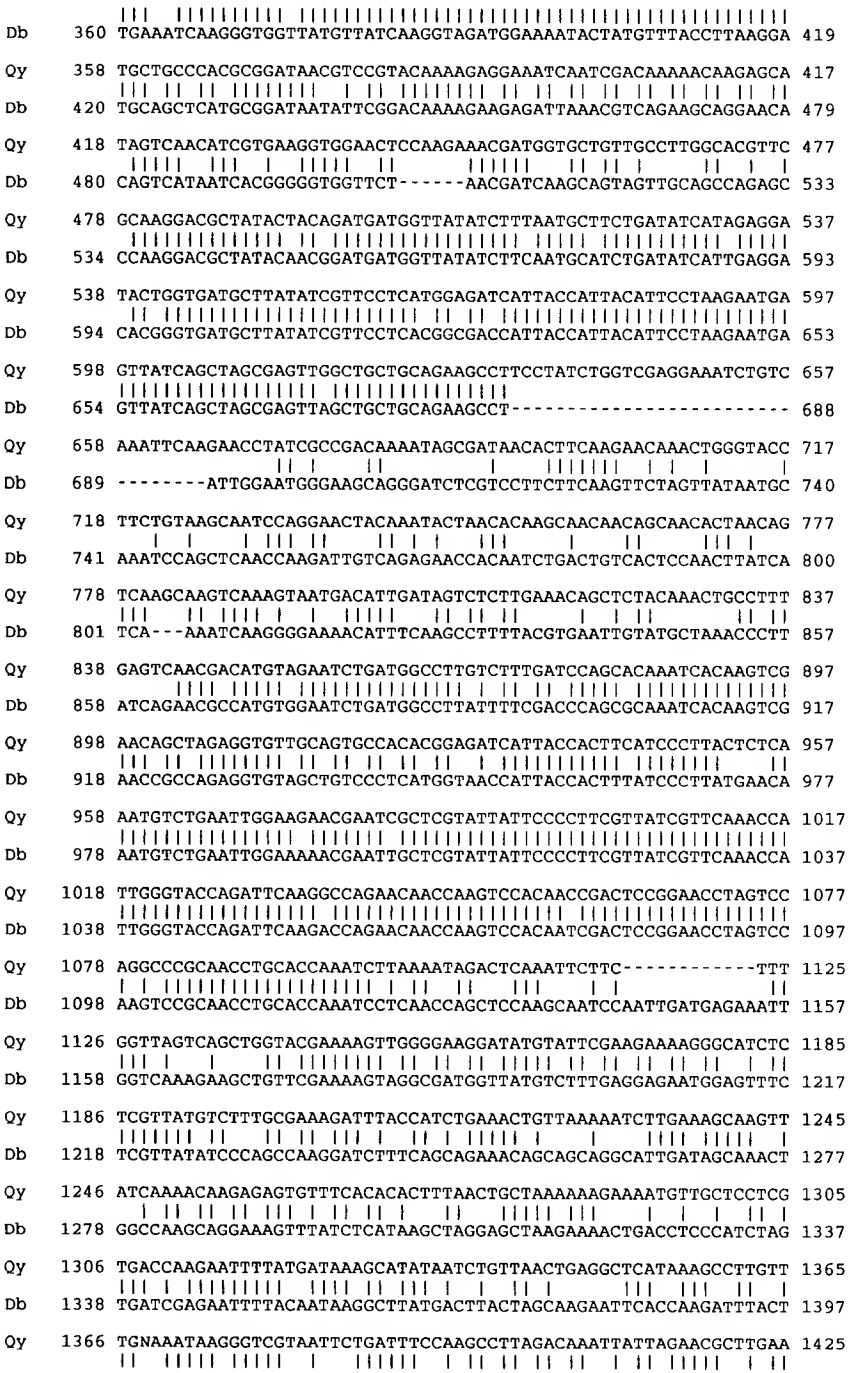
```
Qy      1 TTCTTACGAGTTGGGACTGTATCAAGCTAGAACGGTTAAGGAAAA--TAATCGTGTTTC 57
Db     60 TTCCTATGAGCTTGGACGTTACCAAGCTGGTCAGGATAAGAAAGAGTCTAATCGAGTTGC 119

Qy     58 CTATATAGATGGAAAACAAGCGACGCAAAAAACGGAGAATTTGACTCCTGATGAGGTTAG 117
Db    120 TTATATAGATGGTGATCAGGCTGGTCAAAAGGCAGAAAACTTGACACCAGATGAAGTCAG 179

Qy    118 CAAGCGTGAAGGAATCAATGCTGAGCAAATCGTCATCAAGATAACAGACCAAGGCTATGT 177
Db    180 TAAGAGGGAGGGGATCAACGCCGAACAAATTGTTATCAAGATTACGGATCAAGGTTATGT 239

Qy    178 CACTTCACATGGCGACCACTATCATTATTACAATGGTAAGGTTCCTTATGACGCTATCAT 237
Db    240 GACCTCTCATGGAGACCATTATCATTACTATAATGGCAAGGTTCCTTATGATGCCATCAT 299

Qy    238 CAGTGAAGAATTACTCATGAAAGATCCAAACTATAAGCTAAAAGATGAGGATATTGTTAA 297
Db    300 CAGTGAAGAGCTCCTCATGAAAGATCCGAATTATCAGTTGAAGGATTCAGACATTGTCAA 359

Qy    298 TGAGGTCAAGGGTGGATATGTTATCAAGGTAGATGGAAAATACTATGTTTACCTTAAGGA 357
Db    360 TGAAATCAAGGGTGGTTATGTCATTAAGGTAAACGGTAAATACTATGTTTACCTTAAGGA 419

Qy    358 TGCTGCCCACGCGGATAACGTCCGTACAAAAGAGGAAATCAATCGACAAAAACAAGAGCA 417
Db    420 TGCRGCTCATGCGGATAATATTCGGACAAAAGAAGAGATTAAACGTCAGAAGCAGGAACG 479

Qy    418 TAGTCAACATCGTGAAGGTGGAACTCCAAGAAACGATGGTGCTGTTGCCTTGGCACGTTC 477
Db    480 CAGTCATAATCAT---------AACTCAAGAGCAGATAATGCTGTTGCTGCAGCCAGAGC 530

Qy    478 GCAAGGACGCTATACTACAGATGATGGTTATATCTTTAATGCTTCTGATATCATAGAGGA 537
Db    531 CCAAGGACGTTATACAACGGATGATGGGTATATCTTCAATGCATCTGATATCATTGAGGA 590

Qy    538 TACTGGTGATGCTTATATCGTTCCTCATGGAGATCATTACCATTACATTCCTAAGAATGA 597
Db    591 CACGGGTGATGCTTATATCGTTCCTCACGGCGACCATTACCATTACATTCCTAAGAATGA 650

Qy    598 GTTATCAGCTAGCGAGTTGGCTGCTGCAGAAGCCTTCCTATCTGGTCGAGGAAATCTGTC 657
Db    651 GTTATCAGCTAGCGAGTTAGCTGCTGCAGAAGCCTATTGGAATGG-------------- 695

Qy    658 AAATTCAAGAACCTATCGCCGACAAAATAGCGATAACACTTCAAGAACAAACTGGGTACC 717
Db    696 -GAAGCAGGGATCTCGTCCTTCTTCAAGTTCTAGTTATAATGCAAATCCAGCTCAACCAA 754

Qy    718 TTCTGTAAGCAATCCAGGAACTACAAATACTAACACAAGCAACAACAGCAACACTAACAG 777
Db    755 GATTGTCAGAGAACCACAATCT-------------------GACTGTCACTCCAACTTA 794

Qy    778 TCAAGCAAGTCAAAGTAATGACATTGATAGTCTCTTGAAACAGCTCTACAAACTGCCTTT 837
Db    795 TCATCAAAATCAAGGGGAAAACATTTCAAGCCTTTTACGTGAATTGTATGCTAAACCCTT 854

Qy    838 GAGTCAACGACATGTAGAATCTGATGGCCTTGTCTTTGATCCAGCACAAATCACAAGTCG 897
Db    855 ATCAGAACGCCATGTGGAATCTGATGGCCTTATTTTCGACCCAGCGCAAATCACAAGTCG 914

Qy    898 AACAGCTAGAGGTGTTGCAGTGCCACACGGAGATCATTACCACTTCATCCCTTACTCTCA 957
Db    915 AACCGCCAGAGGTGTAGCTGTCCCTCATGGTAACCATTACCACTTTATCCCTTATGAACA 974

Qy    958 AATGTCTGAATTGGAAGAACGAATCGCTCGTATTATTCCCCTTCGTTATCGTTCAAACCA 1017
Db    975 AATGTCTGAATTGGAAAAACGAATTGCTCGTATTATTCCCCTTCGTTATCGTTCAAACCA 1034

Qy   1018 TTGGGTACCAGATTCAAGGCCAGAACAACCAAGTCCACAACCGACTCCGGAACCTAGTCC 1077
Db   1035 TTGGGTACCAGATTCAAGACCAGAAGAACCAAGTCCACAACCGACTCCAGAACCTAGTCC 1094

Qy   1078 AGGCCCGCAACCTGCACCAAATCTTAAAATAGACTCAAATTCTTCTTTGGTTAGTCAGCT 1137
Db   1095 AAGTCCGCAACCAGCTCCAAGCAATCCAATTGATGGGAA------ATTGGTCAAAGAAGC 1148

Qy   1138 GGTACGAAAAGTTGGGGAAGGATATGTATTCGAAGAAAAGGGCATCTCTCGTTATGTCTT 1197
Db   1149 TGTTCGAAAAGTAGGCGATGGTTATGTCTTTGAGGAGAATGGAGTTTCTCGTTATATCCC 1208

Qy   1198 TGCGAAAGATTTACCATCTGAAACTGTTAAAAATCTTGAAAGCAAGTTATCAAAACAAGA 1257
Db   1209 AGCCAAGGATCTTTCAGCAGAAACAGCAGCAGGCATTGATAGCAAACTGGCCAAGCAGGA 1268

Qy   1258 GAGTGTTTCACACACTTTAACTGCTAAAAAAGAAAATGTTGCTCCTCGTGACCAAGAATT 1317
Db   1269 AAGTTTATCTCATAAGCTAGGAACTAAGAAAACTGACCTCCCATCTAGTGATCGAGAATT 1328

Qy   1318 TTATGATAAAGCATATAATCTGTTAACTGAGGCTCATAAAGCCTTGTTTGNAAATAAGGG 1377
Db   1329 TTACAATAAGGCTTATGACTTACTAGCAAGAATTCACCAAGATTTACTTGATAATAAAGG 1388

Qy   1378 TCGTAATTCTGATTTCCAAGCCTTAGACAAATTATTAGAACGCTTGAATGATGAATCGAC 1437
Db   1389 TCGACAAGTTGATTTTGAGGCTTTGGATAACCTGTTGGAACGACTCAAGGATGTCTCAAG 1448

Qy   1438 TAATAAAGAAAAATTGGTAGATGATTTATTGGCATTCCTAGCACCAATTACCCATCCAGA 1497
Db   1449 TGATAAAGTCAAGTTAGTGGAAGATATTCTTGCCTTCTTAGCTCCGATTCGTCATCCAGA 1508

Qy   1498 GCGACTTGGCAAACCAAATTCTCAAATTGAGTATACTGAAGACGAAGTTCGTATTGCTCA 1557
Db   1509 ACGTTTAGGAAAACCAAATGCGCAAATTACCTACACTGATGATGAGATTCAAGTAGCCAA 1568

Qy   1558 ATTAGCTGATAAGTATACAACGTCAGATGGTTACATTTTTGATGAACATGATATAATCAG 1617
Db   1569 GTTGGCAGGCAAGTACACAGCAGAAGACGGTTATATCTTTGATCCTCGTGATATAACCAG 1628

Qy   1618 TGATGAAGGAGATGCATATGTAACGCCTCATATGGGCCATAGTCACTGGATTGGAAAAGA 1677
Db   1629 TGATGAGGGGGATGCCTATGTAACTCCACATATGACCCATAGCCACTGGATTAAAAAAGA 1688

Qy   1678 TAGCCTTTCTGATAAGGAAAAAGTTGCAGCTCAAGCCTATACTAAAGAAAAAGGTATCCT 1737
Db   1689 TAGTTTGTCTGAAGCTGAGAGAGCGGCAGCCCAGGCTTATGCTRAAGAGAAAGGTTTGAC 1748

Qy   1738 ACCTCCATCTCCAGACGCAGATGTTAAAGCAAATCCAACTGGAGATAGTGCAGCAGCTAT 1797
Db   1749 CCCTCCTTCGACAGACCATCAGGATTCAGGAAATACTGAGGCAAAAGGAGCAGAAGCTAT 1808

Qy   1798 TTACAATCGTGTGAAAGGGGAAAAACGAATTCCACTCGTTCGACTTCCATATATGGTTGA 1857
Db   1809 CTACAACCGMGTGAAAGCAGCTAAGAAGGTGCCACTTGATCGTATGCCTTACAATCTTCA 1868

Qy   1858 GCATACAGTTGAGGTTAAAAACGGTAATTTGATTATTCCTCATAAGGATCATTACCATAA 1917
Db   1869 ATATACTGTAGAAGTCAAAAACGGTAGTTTAATCATACCTCATTATGACCATTACCATAA 1928

Qy   1918 TATTAAATTTGCTTGGTTTGATGATCACACATACAAAGCTCCAAATGGCTATACCTTGGA 1977
Db   1929 CATCAAATTTGAGTGGTTTGACGAAGGCCTTTATGAGGCACCTAAGGGGTATACTCTTGA 1988
```

```
Qy   1978 AGATTTGTTTGCGACGATTAAGTACTACGTAGAACACCCTGACGAACGTCCACATTCTAA 2037
          ||| | || |||||  | ||||||||| || ||||| ||  ||||||||||| |||||  |
Db   1989 GGATCTTTTGGCGACTGTCAAGTACTATGTCGAACATCCAAACGAACGTCCGCATTCAGA 2048

Qy   2038 TGATGGATGGGGCAATGCCAGTGAGCATGTGTTAGGCAAGAAAGACCACAGTGAAGATCC 2097
          | |||| |  || || || || || ||||||   | | || ||| |       |  ||| |
Db   2049 TAATGGTTTTGGTAACGCTAGCGACCATGTTCAAAGAAACAAAAATGGTCAAGCTGATAC 2108

Qy   2098 AAATAAGAACTTCAAAGCGGATGA---------------------------------- 2121
          ||| | |     ||| |    ||
Db   2109 CAATCAAACGGAAAAACCAAGCGAGGAGAAACCTCAGACAGAAAAACCTGAGGAAGAAAC 2168

Qy   2122 ------------------------------------------------AGAGCCAGT 2130
                                                          ||||  |
Db   2169 CCCTCGAGAAGAGAAACCGCAAAGCGAGAAACCAGAGTCTCCAAAACCAACAGAGGAACC 2228

Qy   2131 AGAGGAAACACCTGCTGAGCCAGAAGTCCCTCAAGTAGAGACTGAAAAAGTAGAAGCCCA 2190
          ||| ||| |||| |  ||  |||||| |||||| || |||||||||||| || ||||  |
Db   2229 AGAAGAATCACCAGAGGAATCAGAAGAACCTCAGGTCGAGACTGAAAAGGTTGAAGAAAA 2288

Qy   2191 ACTCAAAGAAGCAGAAGTTTTGCTTGCGAAAGTAACGGATTCTAGTCTGAAAGCCAATGC 2250
          ||| | ||| || |||| ||| |||  ||| |   |||| | | | | ||  |||||||
Db   2289 ACTGAGAGAGGCTGAAGATTTACTTGGAAAAATCCAGGATCCAATTATCAAGTCCAATGC 2348

Qy   2251 AACAGAAACTCTAGCTGGTTTACGAAATAATTTGACTCTTCAAATTATGGATAACAATAG 2310
          | ||| |||||  | || |||  |||||||||     ||   |    ||| |||||||
Db   2349 CAAAGAGACTCTCACAGGATTAAAAAATAATTTACTATTTGGCACCCAGGACAACAATAC 2408

Qy   2311 TATCATGGCAGAAGCAGAAAAATTACTTGCGTTGTTAAAAGGAAGTAA 2358
          ||| ||||||||||| |||||| || | || || ||||| |  |||||
Db   2409 TATTATGGCAGAAGCTGAAAAACTATTGGCTTTATTAAAGGAGAGTAA 2456
```

RESULT 11
AF318955
LOCUS AF318955 2517 bp DNA linear BCT 11-FEB-2001
DEFINITION Streptococcus pneumoniae pneumococcal histidine triad protein D precursor (phtD) gene, partial cds.
ACCESSION AF318955
VERSION AF318955.1 GI:12744743
KEYWORDS .
SOURCE Streptococcus pneumoniae.
ORGANISM Streptococcus pneumoniae
Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Streptococcus.
REFERENCE 1 (bases 1 to 2517)
AUTHORS Adamou,J.E., Heinrichs,J.H., Erwin,A.L., Walsh,W., Gayle,T., Dormitzer,M., Dagan,R., Brewah,Y.A., Barren,P., Lathigra,R., Langermann,S., Koenig,S. and Johnson,S.
TITLE Identification and characterization of a novel family of pneumococcal proteins that are protective against sepsis
JOURNAL Infect. Immun. 69 (2), 949-958 (2001)
MEDLINE 21101045
PUBMED 11159990
REFERENCE 2 (bases 1 to 2517)
AUTHORS Adamou,J.E., Heinrichs,J.H., Erwin,A.L., Walsh,W., Dormitzer,M. and Johnson,S.
TITLE Direct Submission
JOURNAL Submitted (03-NOV-2000) Molecular Microbiology, MedImmune, Inc., 35 West Watkins Mill Road, Gaithersburg, MD 20878, USA
FEATURES Location/Qualifiers
source 1. .2517
/organism="Streptococcus pneumoniae"
/serotype="4"
/db_xref="taxon:1313"
gene 1. .>2517
/gene="phtD"
CDS 1. .>2517
/gene="phtD"
/note="PhtD; protective pneumococcal antigen"
/codon_start=1
/transl_table=11
/product="pneumococcal histidine triad protein D precursor"
/protein_id="AAK06760.1"
/db_xref="GI:12744744"
/translation="MKINKKYLAGSVAVLALSVCSYELGRHQAGQVKKESNRVSYIDG
DQAGQKAENLTPDEVSKREGINAEQIVIKITDQGYVTSHGDHYHYYNGKVPYDAIISE
ELLMKDPNYQLKDSDIVNEIKGGYVIKVDGKYYVYLKDAAHADNIRTKEEIKRQKQEH
SHNHGGGSNDQAVVAARAQGRYTTDDGYIFNASDIIEDTGDAYIVPHGDHYHYIPKNE
LSASELAAAEAYWNGKQGSRPSSSSSYNANPAQPRLSENHNLTVTPTYHQNQGENISS
LLRELYAKPLSERHVESDGLIFDPAQITSRTARGVAVPHGNHYHFIPYEQMSELEKRI
ARIIPLRYRSNHWVPDSRPEQPSPQSTPEPSPSPQPAPNPQPAPSNPIDEKLVKEAVR
KVGDGYVFEENGVSRYIPAKDLSAETAAGIDSKLAKQESLSHKLGAKKTDLPSSDREF
YNKAYDLLARIHQDLLDNKGRQVDFEALDNLLERLKDVPSDKVKLVDDILAFLAPIRH
PERLGKPNAQITYTDDEIQVAKLAGKYTTEDGYIFDPRDITSDEGDAYVTPHMTHSHW
IKKDSLSEAERAAAQAYAKEKGLTPPSTDHQDSGNTEAKGAEAIYNRVKAAKKVPLDR
MPYNLQYTVEVKNGSLIIPHYDHYHNIKFEWFDEGLYEAPKGYTLEDLLATVKYYVEH
PNERPHSDNGFGNASDHVRKNKVDQDSKPDEDKEHDEVSEPTHPESDEKENHAGLNPS
ADNLYKPSTDTEETEEEAEDTTDEAEIPQVENSVINAKIADAEALLEKVTDPSIRQNA
METLTGLKSSLLLGTKDNNTISAEVDSLLALLKESQPAPIQ"
sig_peptide 1. .87
/gene="phtD"
misc_feature 247. .264
/gene="phtD"
/note="Region: histidine triad"
misc_feature 418. .495
/gene="phtD"
/note="Region: coiled-coil domain"
misc_feature 619. .636
/gene="phtD"
/note="Region: histidine triad"
misc_feature 943. .960
/gene="phtD"
/note="Region: histidine triad"
misc_feature 1417. .1509
/gene="phtD"
/note="Region: coiled-coil domain"
misc_feature 1678. .1695
/gene="phtD"
/note="Region: histidine triad"
misc_feature 1930. .1947
/gene="phtD"
/note="Region: histidine triad"
repeat_region 526. .717
/rpt_type=direct
repeat_region 1597. .1779
/rpt_type=direct
BASE COUNT 830 a 509 c 545 g 633 t
ORIGIN

Query Match 41.4%; Score 990; DB 1; Length 2517;
Best Local Similarity 68.1%; Pred. No. 3.2e-201;
Matches 1478; Conservative 0; Mismatches 636; Indels 57; Gaps 5;

```
Qy      1 TTCTTACGAGTTGGGACTGTATCAAGCTAGAACGGTTAAGGAAAA---TAATCGTGTTTC 57
          ||| || ||  | || |   | |||||| |   ||||||| || |   |||||| |||||
Db     60 TTCCTATGAACTTGGTCGTCACCAAGCTGGTCAGGTTAAGAAAGAGTCTAATCGAGTTTC 119

Qy     58 CTATATAGATGGAAAACAAGCGACGCAAAAAACGGAGAATTTGACTCCTGATGAGGTTAG 117
          ||||||||||  | || ||    |||||  | || || ||||| || ||||| || ||
Db    120 TTATATAGATGGTGATCAGGCTGGTCAAAAGGCAGAAAACTTGACACCAGATGAAGTCAG 179

Qy    118 CAAGCGTGAAGGAATCAATGCTGAGCAAATCGTCATCAAGATAACAGACCAAGGCTATGT 177
          ||| | || || ||||| || || |||||||||||||||||| || || ||||| |||||
Db    180 TAAGAGGGAGGGGATCAACGCCGAACAAATCGTCATCAAGATTACGGATCAAGGTTATGT 239

Qy    178 CACTTCACATGGCGACCACTATCATTATTACAATGGTAAGGTTCCTTATGACGCTATCAT 237
          || || ||||| ||||| |||||||| || ||||| ||||| |||||||| || |||||
Db    240 GACCTCTCATGGAGACCATTATCATTACTATAATGGCAAGGTCCCTTATGATGCCATCAT 299

Qy    238 CAGTGAAGAATTACTCATGAAAGATCCAAACTATAAGCTAAAAGATGAGGATATTGTTAA 297
          |||||||||  | |||||||||||||| || ||| || | || |||   || ||||| ||
Db    300 CAGTGAAGAGCTCCTCATGAAAGATCCGAATTATCAGTTGAAGGATTCAGACATTGTCAA 359

Qy    298 TGAGGTCAAGGGTGGATATGTTATCAAGGTAGATGGAAAATACTATGTTTACCTTAAGGA 357
```



RESULT 12
AE007403
LOCUS AE007403 11931 bp DNA linear BCT 31-AUG-2001
DEFINITION Streptococcus pneumoniae TIGR4 section 86 of 194 of the complete genome.
ACCESSION AE007403 AE005672
VERSION AE007403.1 GI:14972469
KEYWORDS .
SOURCE Streptococcus pneumoniae TIGR4.
ORGANISM Streptococcus pneumoniae TIGR4
Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Streptococcus.
REFERENCE 1 (bases 1 to 11931)
AUTHORS Tettelin,H., Nelson,K.E., Paulsen,I.T., Eisen,J.A., Read,T.D., Peterson,S., Heidelberg,J., DeBoy,R.T., Haft,D.H., Dodson,R.J., Durkin,A.S., Gwinn,M., Kolonay,J.F., Nelson,W.C., Peterson,J.D., Umayam,L.A., White,O., Salzberg,S.L., Lewis,M.R., Radune,D., Holtzapple,E., Khouri,H., Wolf,A.M., Utterback,T.R., Hansen,C.L.,

```
            McDonald,L.A., Feldblyum,T.V., Angiuoli,S., Dickinson,T.,
            Hickey,E.K., Holt,I.E., Loftus,B.J., Yang,F., Smith,H.O.,
            Venter,J.C., Dougherty,B.A., Morrison,D.A., Hollingshead,S.K. and
            Fraser,C.M.
  TITLE     Complete genome sequence of a virulent isolate of Streptococcus
            pneumoniae
  JOURNAL   Science 293 (5529), 498-506 (2001)
  MEDLINE   21357209
   PUBMED   11463916
REFERENCE   2  (bases 1 to 11931)
  AUTHORS   Tettelin,H., Nelson,K.E., Paulsen,I.T., Eisen,J.A., Read,T.D.,
            Peterson,S., Heidelberg,J., DeBoy,R.T., Haft,D.H., Dodson,R.J.,
            Durkin,A.S., Gwinn,M., Kolonay,J.F., Nelson,W.C., Peterson,J.D.,
            Umayam,L.A., White,O., Lewis,M.R., Radune,D., Holtzapple,E.,
            Khouri,H., Wolf,A.M., Utterback,T.R., Hansen,C.L., McDonald,L.A.,
            Feldblyum,T.V., Angiuoli,S., Gesuwan,P., Hickey,E.K., Holt,I.E.,
            Loftus,B.J., Ujwal,M.L., Yang,F., Smith,H.O., Venter,J.C.,
            Dougherty,B.A., Morrison,D.A., Hollingshead,S.K. and Fraser,C.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-JUN-2001) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA
FEATURES             Location/Qualifiers
     source          1. .11931
                     /organism="Streptococcus pneumoniae TIGR4"
                     /strain="TIGR4"
                     /db_xref="taxon:170187"
     gene            129. .476
                     /gene="SP0995"
     gene            655. .795
                     /gene="SP0996"
                     /note="This region contains a match to at least one other
                     gene that is not full length,and is not the result of a
                     sequencing artifact; similar to GP:5019553; identified by
                     sequence similarity; putative; IS630-Spn1, transposase
                     Orf2, truncation"
     gene            956. .1156
                     /gene="SP0997"
     CDS             956. .1156
                     /gene="SP0997"
                     /note="identified by Glimmer2; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAK75114.1"
                     /db_xref="GI:14972470"
                     /translation="MVASASASSTSTQAQEQVDKSELRALSQELDQRLKALATVSDPK
                     IDATKAVLLDAQKAPEDSALTE"
     gene            1303. .1602
                     /gene="SP0998"
     CDS             1303. .1602
                     /gene="SP0998"
                     /note="identified by Glimmer2; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAK75115.1"
                     /db_xref="GI:14972471"
                     /translation="MTETNSVPAGVIVVSLLALLGVIAFWLIRRKKESEIQQLSTELI
                     KVLGQLDAEKADKKVLAKAQNLLQETLDFVKEENGSAETETKLVEELKAILDKLK"
     gene            1833. .2540
                     /gene="SP0999"
     CDS             1833. .2540
                     /gene="SP0999"
                     /note="similar to GP:6714962; identified by sequence
                     similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="cytochrome c-type biogenesis protein CcdA"
                     /protein_id="AAK75116.1"
                     /db_xref="GI:14972472"
                     /translation="METIVFLISVFLAGVLSFFSPCIFPLLPVYAGILLDDQESAKSF
                     SLFGRKVLWSGLIRTLCFIAGISLIFFILGFGAGYFGHILYANWFRYGMGAIIILGL
                     HQMEIFHLKKLEVQKSFTFKKSDSNRYWSAFLLGITFSFGWTPCIGPVLSSVLALAAS
                     GGNGAWQGAIYTLIYTLGMALPFLVLALASGLVMPYFSKIKRHMMLLKKIGGFLIVLM
                     GILLLLGQVNVLAGIFE"
     gene            2551. .3108
                     /gene="SP1000"
     CDS             2551. .3108
                     /gene="SP1000"
                     /codon_start=1
                     /transl_table=11
                     /product="thioredoxin family protein"
                     /protein_id="AAK75117.1"
                     /db_xref="GI:14972473"
                     /translation="MKKVMFAGLSLLSLVVLMACGEEETKKTQAAQQPKQQTTVQQIA
                     VGKDAPDFTLQSMDGKEVKLSDFKGKKVYLKFWASWCGPCKKSMPELMELAAKPDRDF
                     EILTVIAPGIQGEKTVEQFPQWFQEQGYKDIPVLYDTKATTFQAYQIRSIPTEYLIDS
                     QGKIGKIQFGAISNADAEAAFKEMN"
     gene            3482. .4873
                     /gene="SP1001"
     CDS             3482. .4873
                     /gene="SP1001"
                     /note="identified by match to PFAM  protein family HMM
                     PF00324"
                     /codon_start=1
                     /transl_table=11
                     /product="amino acid permease family protein"
                     /protein_id="AAK75118.1"
                     /db_xref="GI:14972474"
                     /translation="MNIFRTKNVSLDKTEMHRHLKLWDLILLGIGAMVGTGVFTITGT
                     AAATLAGPALVISIVISALCVGLSALFFAEFASRVPATGGAYSYLYAILGEFPAWLAG
                     WLTMMEFMTAISGVASGWAAYFKGLLSQYGIAFPQALNGTFNPQAGTFVDLLPILVLV
                     LVTSLVLLNAKAALRFNSILVILKFSALALFVLVGIWNIKFDNWSNFAPYGFGQIYGA
                     STGIMAGASLMFFGFLGFESISMAVDEVKTPQKNIPRGIVLSLSIVTILYALVTLVLT
                     GVVHYSHLNVDDAVAFALRSVGISWAANYVSLVAILTLITVCISMTYALSRMIYSLAR
                     DGLVPAAFKELTKTSKIPKNATILTGLASAVAAGMFPLASIAAFLNICTLAYLIMLAY
                     GLIRLRKEKGMPKAGEFKTPLVPLLPILSIIICLSFMLQYNMNTWIAFLVALLVGSII
                     YFTYGYKHSTIEE"
     gene            5016. .5933
                     /gene="SP1002"
     CDS             5016. .5933
                     /gene="SP1002"
                     /note="identified by match to PFAM  protein family HMM
                     PF01297"
                     /codon_start=1
                     /transl_table=11
                     /product="adhesion lipoprotein"
                     /protein_id="AAK75119.1"
                     /db_xref="GI:14972475"
                     /translation="MKKQNLFLVLLSVFLLCLGACGQKESQTGKGMKIVTSFYPIYAM
                     VKEVSGDLNDVRMIQSSSGIHSFEPSANDIAAIYDADVFVYHSHTLESWAGSLDPNLK
                     KSKVKVLEASEGMTLERVPGLEDVEAGDGVDEKTLYDPHTWLDPEKAGEEAQIIADKL
                     SEVDSEHKETYQKNAQAFIKKAQELTKKFQPKFEKATQKTFVTQHTAFSYLAKRFGLN
                     QLGIAGISPEQEPSPRQLTEIQEFVKTYKVKTIFTESNASSKVAETLVKSTGVGLKTL
                     NPLESDPQNDKTYLENLEENMSILAEELK"
     gene            5941. .8460
                     /gene="SP1003"
     CDS             5941. .8460
                     /gene="SP1003"
                     /note="identified by Glimmer2; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="conserved hypothetical protein"
                     /protein_id="AAK75120.1"
                     /db_xref="GI:14972476"
                     /translation="MKINKKYLAGSVAVLALSVCSYELGRHQAGQVKKESNRVSYIDG
                     DQAGQKAENLTPDEVSKREGINAEQIVIKITDQGYVTSHGDHYHYYNGKVPYDAIISE
                     ELLMKDPNYQLKDSDIVNEIKGGYVIKVDGKYYVYLKDAAHADNIRTKEEIKRQKQEH
                     SHNHGGGSNDQAVVAARAQGRYTTDDGYIFNASDIIEDTGDAYIVPHGDHYHYIPKNE
                     LSASELAAAEAYWNGKQGSRPSSSSSYNANPAQPRLSENHNLTVTPTYHQNQGENISS
                     LLRELYAKPLSERHVESDGLIFDPAQITSRTARGVAVPHGNHYHFIPYEQMSELEKRI
                     ARIIPLRYRSNHWVPDSRPEQPSPQSTPEPSPSPQPAPNPQPAPSNPIDEKLVKEAVR
                     KVGDGYVFEENGVSRYIPAKDLSAETAAGIDSKLAKQESLSHKLGAKKTDLPSSDREF
                     YNKAYDLLARIHQDLLDNKGRQVDFEALDNLLERLKDVPSDKVKLVDDILAFLAPIRH
                     PERLGKPNAQITYTDDEIQVAKLAGKYTTEDGYIFDPRDITSDEGDAYVTPHMTHSHW
                     IKKDSLSEAERAAAQAYAKEKGLTPPSTDHQDSGNTEAKGAEAIYNRVKAAKKVPLDR
```

```
                     MPYNLQYTVEVKNGSLIIPHYDHYHNIKFEWFDEGLYEAPKGYTLEDLLATVKYYVEH
                     PNERPHSDNGFGNASDHVRKNKVDQDSKPDEDKEHDEVSEPTHPESDEKENHAGLNPS
                     ADNLYKPSTDTEETEEEAEDTTDEAEIPQVENSVINAKIADAEALLEKVTDPSIRQNA
                     METLTGLKSSLLLGTKDNNTISAEVDSLLALLKESQPAPIQ"
     gene            8670. .11789
                     /gene="SP1004"
     CDS             8670. .11789
                     /gene="SP1004"
                     /note="identified by Glimmer2; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="conserved hypothetical protein"
                     /protein_id="AAK75121.1"
                     /db_xref="GI:14972477"
                     /translation="MKFSKKYIAAGSAVIVSLSLCAYALNQHRSQENKDNNRVSYVDG
                     SQSSQKSENLTPDQVSQKEGIQAEQIVIKITDQGYVTSHGDHYHYYNGKVPYDALFSE
                     ELLMKDPNYQLKDADIVNEVKGGYIIKVDGKYYVYLKDAAHADNVRTKDEINRQKQEH
                     VKDNEKVNSNVAVARSQGRYTTNDGYVFNPADIIEDTGNAYIVPHGGHYHYIPKSDLS
                     ASELAAAKAHLAGKNMQPSQLSYSSTASDNNTQSVAKGSTSKPANKSENLQSLLKELY
                     DSPSAQRYSESDGLVFDPAKIISRTPNGVAIPHGDHYHFIPYSKLSALEEKIARMVPI
                     SGTGSTVSTNAKPNEVVSSLGSLSSNPSSLTTSKELSSASDGYIFNPKDIVEETATAY
                     IVRHGDHFHYIPKSNQIGQPTLPNNSLATPSPSLPINPGTSHEKHEEDGYGFDANRII
                     AEDESGFVMSHGDHNHYFFKKDLTEEQIKAAQKHLEEVKTSHNGLDSLSSHEQDYPSN
                     AKEMKDLDKKIEEKIAGIMKQYGVKRESIVVNKEKNAIIYPHGDHHHADPIDEHKPVG
                     IGHSHSNYELFKPEEGVAKKEGNKVYTGEELTNVVNLLKNSTFNNQNFTLANGQKRVS
                     FSFPPELEKKLGINMLVKLITPDGKVLEKVSGKVFGEGVGNIANFELDQPYLPGQTFK
                     YTIASKDYPEVSYDGTFTVPTSLAYKMASQTIFYPFHAGDTYLRVNPQFAVPKGTDAL
                     VRVFDEFHGNAYLENNYKVGEIKLPIPKLNQGTTRTAGNKIPVTFMANAYLDNQSTYI
                     VEVPILEKENQTDKPSILPQFKRNKAQENLKLDEKVEEPKTSEKVEKEKLSETGNSTS
                     NSTLEEVPTVDPVQEKVAKFAESYGMKLENVLFNMDGTIELYLPSGEVIKKNMADFTG
                     EAPQGNGENKPSENGKVSTGTVENQPTENKPADSLPEAPNEKPVKPENSTDNGMLNPE
                     GNVGSDPMLDPALEEAPAVDPVQEKLEKFTASYGLGLDSVIFNMDGTIELRLPSGEVI
                     KKNLSDLIA"
BASE COUNT     3963 a    2120 c    2445 g    3403 t
ORIGIN

  Query Match             41.4%;  Score 990;  DB 1;  Length 11931;
  Best Local Similarity   68.1%;  Pred. No. 3.1e-201;
  Matches 1478;  Conservative    0;  Mismatches  636;  Indels   57;  Gaps    5;

Qy      1 TTCTTACGAGTTGGGACTGTATCAAGCTAGAACGGTTAAGGAAAA---TAATCGTGTTTC 57
          ||| || ||  | || |   | ||||||  |   |||||||  || |   ||||||  |||||
Db   6000 TTCCTATGAACTTGGTCGTCACCAAGCTGGTCAGGTTAAGAAAGAGTCTAATCGAGTTTC 6059

Qy     58 CTATATAGATGGAAAACAAGCGACGCAAAAAACGGAGAATTTGACTCCTGATGAGGTTAG 117
          ||||||||||||  | || ||    |||||  | || || ||||| || ||||| || ||
Db   6060 TTATATAGATGGTGATCAGGCTGGTCAAAAGGCAGAAAACTTGACACCAGATGAAGTCAG 6119

Qy    118 CAAGCGTGAAGGAATCAATGCTGAGCAAATCGTCATCAAGATAACAGACCAAGGCTATGT 177
          ||| | || || ||||| || || ||||||||||||||||| || || ||||| |||||
Db   6120 TAAGAGGGAGGGGATCAACGCCGAACAAATCGTCATCAAGATTACGGATCAAGGTTATGT 6179

Qy    178 CACTTCACATGGCGACCACTATCATTATTACAATGGTAAGGTTCCTTATGACGCTATCAT 237
          || || ||||| ||||| |||||||| || ||||| ||||| |||||||| || |||||
Db   6180 GACCTCTCATGGAGACCATTATCATTACTATAATGGCAAGGTCCCTTATGATGCCATCAT 6239

Qy    238 CAGTGAAGAATTACTCATGAAAGATCCAAACTATAAGCTAAAAGATGAGGATATTGTTAA 297
          |||||||||  | ||||||||||||||| || ||| || | || |||   || |||||
Db   6240 CAGTGAAGAGCTCCTCATGAAAGATCCGAATTATCAGTTGAAGGATTCAGACATTGTCAA 6299

Qy    298 TGAGGTCAAGGGTGGATATGTTATCAAGGTAGATGGAAAATACTATGTTTACCTTAAGGA 357
          |||  |||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db   6300 TGAAATCAAGGGTGGTTATGTTATCAAGGTAGATGGAAAATACTATGTTTACCTTAAGGA 6359

Qy    358 TGCTGCCCACGCGGATAACGTCCGTACAAAAGAGGAAATCAATCGACAAAAACAAGAGCA 417
          ||| || || ||||||||  | || |||||||| || || || || || || || || ||
Db   6360 TGCAGCTCATGCGGATAATATTCGGACAAAAGAAGAGATTAAACGTCAGAAGCAGGAACA 6419

Qy    418 TAGTCAACATCGTGAAGGTGGAACTCCAAGAAACGATGGTGCTGTTGCCTTGGCACGTTC 477
          |||||  |||  |  |||||  ||      ||||||   || || |     ||  |  |
Db   6420 CAGTCATAATCACGGGGGTGGTTCT------AACGATCAAGCAGTAGTTGCAGCCAGAGC 6473

Qy    478 GCAAGGACGCTATACTACAGATGATGGTTATATCTTTAATGCTTCTGATATCATAGAGGA 537
          |||||||||||||| || ||||||||||||||| ||||| ||||||||||| |||||
Db   6474 CCAAGGACGCTATACAACGGATGATGGTTATATCTTCAATGCATCTGATATCATTGAGGA 6533

Qy    538 TACTGGTGATGCTTATATCGTTCCTCATGGAGATCATTACCATTACATTCCTAAGAATGA 597
          || |||||||||||||||||||||||| || || |||||||||||||||||||||||||||
Db   6534 CACGGGTGATGCTTATATCGTTCCTCACGGCGACCATTACCATTACATTCCTAAGAATGA 6593

Qy    598 GTTATCAGCTAGCGAGTTGGCTGCTGCAGAAGCCTTCCTATCTGGTCGAGGAAATCTGTC 657
          ||||||||||||||||| ||||||||||||||||
Db   6594 GTTATCAGCTAGCGAGTTAGCTGCTGCAGAAGCCT------------------------- 6628

Qy    658 AAATTCAAGAACCTATCGCCGACAAAATAGCGATAACACTTCAAGAACAAACTGGGTACC 717
                      || |    ||         |    |||||||  | |  |      |
Db   6629 --------ATTGGAATGGGAAGCAGGGATCTCGTCCTTCTTCAAGTTCTAGTTATAATGC 6680

Qy    718 TTCTGTAAGCAATCCAGGAACTACAAATACTAACACAAGCAACAACAGCAACACTAACAG 777
            |  |    | ||| ||    || | |   |||      |    ||     ||| |
Db   6681 AAATCCAGCTCAACCAAGATTGTCAGAGAACCACAATCTGACTGTCACTCCAACTTATCA 6740

Qy    778 TCAAGCAAGTCAAAGTAATGACATTGATAGTCTCTTGAAACAGCTCTACAAACTGCCTTT 837
          |||   || |||| |  |  |||||   || || ||      |  | ||        || ||
Db   6741 TCA---AAATCAAGGGGAAAACATTTCAAGCCTTTTACGTGAATTGTATGCTAAACCCTT 6797

Qy    838 GAGTCAACGACATGTAGAATCTGATGGCCTTGTCTTTGATCCAGCACAAATCACAAGTCG 897
              |||| ||||| ||||||||||||||| | || || ||||| ||||||||||||||
Db   6798 ATCAGAACGCCATGTGGAATCTGATGGCCTTATTTTCGACCCAGCGCAAATCACAAGTCG 6857

Qy    898 AACAGCTAGAGGTGTTGCAGTGCCACACGGAGATCATTACCACTTCATCCCTTACTCTCA 957
          ||| || |||||||| || || || || |  |||||||||| ||||||||    ||
Db   6858 AACCGCCAGAGGTGTAGCTGTCCCTCATGGTAACCATTACCACTTTATCCCTTATGAACA 6917

Qy    958 AATGTCTGAATTGGAAGAACGAATCGCTCGTATTATTCCCCTTCGTTATCGTTCAAACCA 1017
          |||||||||||||||| |||||| ||||||||||||||||||||||||||||||||||||
Db   6918 AATGTCTGAATTGGAAAAACGAATTGCTCGTATTATTCCCCTTCGTTATCGTTCAAACCA 6977

Qy   1018 TTGGGTACCAGATTCAAGGCCAGAACAACCAAGTCCACAACCGACTCCGGAACCTAGTCC 1077
          |||||||||||||||||| ||||||||||||||||||| ||||||||||||||||||||
Db   6978 TTGGGTACCAGATTCAAGACCAGAACAACCAAGTCCACAATCGACTCCGGAACCTAGTCC 7037

Qy   1078 AGGCCCGCAACCTGCACCAAATCTTAAAATAGACTCAAATTCTTC------------TTT 1125
          | | |||||||||||||||||| | || ||  |||  | |              ||
Db   7038 AAGTCCGCAACCTGCACCAAATCCTCAACCAGCTCCAAGCAATCCAATTGATGAGAAATT 7097

Qy   1126 GGTTAGTCAGCTGGTACGAAAAGTTGGGGAAGGATATGTATTCGAAGAAAAGGGCATCTC 1185
          ||| |   |   || |||||||| || || || ||||| || || || || ||  | ||
Db   7098 GGTCAAAGAAGCTGTTCGAAAAGTAGGCGATGGTTATGTCTTTGAGGAGAATGGAGTTTC 7157

Qy   1186 TCGTTATGTCTTTGCGAAAGATTTACCATCTGAAACTGTTAAAAATCTTGAAAGCAAGTT 1245
          |||||||  ||   || || ||| |  || | ||||| |    |    |||| |||||  |
Db   7158 TCGTTATATCCCAGCCAAGGATCTTTCAGCAGAAACAGCAGCAGGCATTGATAGCAAACT 7217

Qy   1246 ATCAAAACAAGAGAGTGTTTCACACACTTTAACTGCTAAAAAAGAAAATGTTGCTCCTCG 1305
            | || || || ||| | || ||  |||| |||     |  |  | || |
Db   7218 GGCCAAGCAGGAAAGTTTATCTCATAAGCTAGGAGCTAAGAAAACTGACCTCCCATCTAG 7277

Qy   1306 TGACCAAGAATTTTATGATAAAGCATATAATCTGTTAACTGAGGCTCATAAAGCCTTGTT 1365
          ||| | |||||||||  |||| || ||| |  |  || |       |||  ||| ||  |
Db   7278 TGATCGAGAATTTTACAATAAGGCTTATGACTTACTAGCAAGAATTCACCAAGATTTACT 7337

Qy   1366 TGNAAATAAGGGTCGTAATTCTGATTTCCAAGCCTTAGACAAATTATTAGAACGCTTGAA 1425
          ||  ||||| |||||  |   ||||||  | || || || ||  | || |||||  | ||
Db   7338 TGATAATAAAGGTCGACAAGTTGATTTTGAGGCTTTGGATAACCTGTTGGAACGACTCAA 7397

Qy   1426 TGATGAATCGACTAATAAAGAAAAATTGGTAGATGATTTATTGGCATTCCTAGCACCAAT 1485
          ||||   | | | ||||||  || || || ||||||  |  | || ||| |||| || ||
Db   7398 GGATGTCCCAAGTGATAAAGTCAAGTTAGTGGATGATATTCTTGCCTTCTTAGCTCCGAT 7457

Qy   1486 TACCCATCCAGAGCGACTTGGCAAACCAAATTCTCAAATTGAGTATACTGAAGACGAAGT 1545
          |   |||||||| ||  | || |||||||| | ||||||   || |||| || ||  |
Db   7458 TCGTCATCCAGAACGTTTAGGAAAACCAAATGCGCAAATTACCTACACTGATGATGAGAT 7517

Qy   1546 TCGTATTGCTCAATTAGCTGATAAGTATACAACGTCAGATGGTTACATTTTTGATGAACA 1605
          ||   | ||  | || || |  |||||  |||||   |||  |||||  || ||||||   |
```

```
Db   7518 TCAAGTAGCCAAGTTGGCAGGCAAGTACACAACAGAAGACGGTTATATCTTTGATCCTCG 7577

Qy   1606 TGATATAATCAGTGATGAAGGAGATGCATATGTAACGCCTCATATGGGCCATAGTCACTG 1665
          |||||||| |||||||||| || ||||| |||||||| || ||||||  |||||| |||||
Db   7578 TGATATAACCAGTGATGAGGGGGATGCCTATGTAACTCCACATATGACCCATAGCCACTG 7637

Qy   1666 GATTGGAAAAGATAGCCTTTCTGATAAGGAAAAAGTTGCAGCTCAAGCCTATACTAAAGA 1725
          ||||  |||||||||  | |||||    || | ||  ||||| || || ||| |||||||
Db   7638 GATTAAAAAAGATAGTTTGTCTGAAGCTGAGAGAGCGGCAGCCCAGGCTTATGCTAAAGA 7697

Qy   1726 AAAAGGTATCCTACCTCCATCTCCAGACGCAGATGTTAAAGCAAATCCAACTGGAGATAG 1785
           |||||| |   .|||| || |||||    | | |  || |||| |    | | |  |
Db   7698 GAAAGGTTTGACCCCTCCTTCGACAGACCATCAGGATTCAGGAAATACTGAGGCAAAAGG 7757

Qy   1786 TGCAGCAGCTATTTACAATCGTGTGAAAGGGGAAAAACGAATTCCACTCGTTCGACTTCC 1845
          |||| |||||| ||||| || |||||||  |  ||     | ||||| | |||  | ||
Db   7758 AGCAGAAGCTATCTACAACCGCGTGAAAGCAGCTAAGAAGGTGCCACTTGATCGTATGCC 7817

Qy   1846 ATATATGGTTGAGCATACAGTTGAGGTTAAAAACGGTAATTTGATTATTCCTCATAAGGA 1905
           || |   || |  |||| || || || |||||||||| ||| || || |||||| | ||
Db   7818 TTACAATCTTCAATATACTGTAGAAGTCAAAAACGGTAGTTTAATCATACCTCATTATGA 7877

Qy   1906 TCATTACCATAATATTAAATTTGCTTGGTTTGATGATCACACATACAAAGCTCCAAATGG 1965
          |||||||||||| || |||||||  |||||||| ||   |   ||  | || || || ||
Db   7878 CCATTACCATAACATCAAATTTGAGTGGTTTGACGAAGGCCTTTATGAGGCACCTAAGGG 7937

Qy   1966 CTATACCTTGGAAGATTTGTTTGCGACGATTAAGTACTACGTAGAACACCCTGACGAACG 2025
          |||||  | || ||| | || |||||  | ||||||| || |||||  ||  |||||||
Db   7938 GTATACTCTTGAGGATCTTTTGGCGACTGTCAAGTACTATGTCGAACATCCAAACGAACG 7997

Qy   2026 TCCACATTCTAATGATGGATGGGGCAATGCCAGTGAGCATGTGTTAGGCAAGAAAGACCA 2085
          ||| |||||  || |||| |  || || || || || |||||      || || |   4
Db   7998 TCCGCATTCAGATAATGGTTTTGGTAACGCTAGCGACCATGTTCGTAAAAATAAGGTAGA 8057

Qy   2086 CAGTGAAGATCCAAATAAGAACTTCAAAGCGGATGAAGAGCCAGTAGAGGAAACACCTGC 2145
          |   ||   |  |  | |  |    || |   |||| ||   |   |||  ||| |   |
Db   8058 CCAAGACAGTAAACCTGATGAAGATAAGGAACATGATGAAGTAAGTGAGCCAACTCACCC 8117

Qy   2146 TGAGCCAGAAG 2156
          |||  | || |
Db   8118 TGAATCTGATG 8128
```

```
RESULT 13
SPNEU1901
LOCUS       SPNEU1901             232807 bp    DNA     linear   HTG 12-JUL-2001
DEFINITION  Streptococcus pneumoniae clone G54, *** SEQUENCING IN PROGRESS ***,
            in ordered pieces.
ACCESSION   AL449923
VERSION     AL449923.1  GI:11545148
KEYWORDS    HTG; HTGS_PHASE2.
SOURCE      Streptococcus pneumoniae.
  ORGANISM  Streptococcus pneumoniae
            Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
            Streptococcus.
REFERENCE   1  (bases 1 to 232807)
  AUTHORS   Dopazo,J., Mendoza,A., Herrero,J., Caldara,F., Humbert,Y.,
            Friedli,L., Guerrier,M., Grand-Schenk,E., Gandin,C., de
            Francesco,M., Polissi,A., Buell,G., Feger,G., Garcia,E., Peitsch,M.
            and Garcia-Bustos,J.F.
  TITLE     Annotated draft genomic sequence from a Streptococcus pneumoniae
            type 19F clinical isolate
  JOURNAL   Microb. Drug Resist. 7 (2), 99-125 (2001)
  MEDLINE   21335329
   PUBMED   11442348
REFERENCE   2  (bases 1 to 232807)
  AUTHORS   Dopazo,J., Mendoza,A., Herrero,J., Caldara,F., Polissi,A.,
            Humbert,Y., Friedli,L., Guerrier,M., Grand-Schenk,E., Gandin,C., de
            Francesco,M., Buell,G., Feger,G., Garcia,E., Peitsch,M. and
            Garcia-Bustos,J.F.
  TITLE     Direct Submission
  JOURNAL   Submitted (31-OCT-2000) Research Department, Glaxo Wellcome, S.A.,
            Severo Ochoa 2, 28760 Tres Cantos, SPAIN
```

```
COMMENT     * NOTE: This is a 'working draft' sequence.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
FEATURES             Location/Qualifiers
     source          1. .232807
                     /organism="Streptococcus pneumoniae"
                     /serotype="19F"
                     /db_xref="taxon:1313"
                     /clone="G54"
BASE COUNT    70984 a  43538 c  47937 g  70253 t     95 others
ORIGIN
```

```
  Query Match             41.4%;  Score 988.4;  DB 2;  Length 232807;
  Best Local Similarity   65.6%;  Pred. No. 6.3e-201;
  Matches 1616;  Conservative    0;  Mismatches  697;  Indels  150;  Gaps    6;

Qy       1 TTCTTACGAGTTGGGACTGTATCAAGCTAGAACGGTTAAGGAAAA---TAATCGTGTTTC 57
           ||| || ||  | || |  || |||||| |   || |||| || |   ||||| ||| |
Db  200352 TTCCTATGAACTTGGGCGTTACCAAGCTGGTCAGGATAAGAAAGAGTCTAATCGAGTTGC 200411

Qy      58 CTATATAGATGGAAAACAAGCGACGCAAAAAACGGAGAATTTGACTCCTGATGAGGTTAG 117
           ||||||||||  | || ||     |||||  | || |||||||| || ||||| || ||
Db  200412 TTATATAGATGGTGATCAGGCTGGTCAAAAGGCAGAAAATTTGACACCAGATGAAGTCAG 200471

Qy     118 CAAGCGTGAAGGAATCAATGCTGAGCAAATCGTCATCAAGATAACAGACCAAGGCTATGT 177
           ||| | || || ||||| || || ||||| || |||||||| || || ||||| |||||
Db  200472 TAAGAGGGAGGGGATCAACGCCGAACAAATTGTTATCAAGATTACGGATCAAGGTTATGT 200531

Qy     178 CACTTCACATGGCGACCACTATCATTATTACAATGGTAAGGTTCCTTATGACGCTATCAT 237
           || || ||||| ||||| |||||||| || ||||| |||||||||||||| || |||||
Db  200532 GACCTCTATGGAGACCATTATCATTACTATAATGGCAAGGTTCCTTATGATGCCATCAT 200591

Qy     238 CAGTGAAGAATTACTCATGAAAGATCCAAACTATAAGCTAAAAGATGAGGATATTGTTAA 297
           |||||||||  | |||||||||||||| || ||| || | || |||   || ||||| ||
Db  200592 CAGTGAAGAGCTCCTCATGAAAGATCCGAATTATCAGTTGAAGGATTCAGACATTGTCAA 200651

Qy     298 TGAGGTCAAGGGTGGATATGTTATCAAGGTAGATGGAAAATACTATGTTTACCTTAAGGA 357
           |||  |||||||||| ||||| || |||||| | || |||||||||||||||||||||||
Db  200652 TGAAATCAAGGGTGGTTATGTCATTAAGGTAAACGGTAAATACTATGTTTACCTTAAGGA 200711

Qy     358 TGCTGCCCACGCGGATAACGTCCGTACAAAAGAGGAAATCAATCGACAAAAACAAGAGCA 417
           ||| || || |||||||| | || |||||||| || || || || || || || || | |
Db  200712 TGCAGCTCATGCGGATAATATTCGGACAAAAGAAGAGATTAAACGTCAGAAGCAGGAACG 200771

Qy     418 TAGTCAACATCGTGAAGGTGGAACTCCAAGAAACGATGGTGCTGTTGCCTTGGCACGTTC 477
           |||||  ||| |       |   |||||  |||  |||||||||    || |  |
Db  200772 CAGTCATAATCAT---------AACTCAAGAGCAGATAATGCTGTTGCTGCAGCCAGAGC 200822

Qy     478 GCAAGGACGCTATACTACAGATGATGGTTATATCTTTAATGCTTCTGATATCATAGAGGA 537
           |||||||||||  || ||||||||||||||||||||||||||||||||||||||||| || ||
Db  200823 CCAAGGACGCTACACCACAGATGATGGTTATATCTTTAATGCTTCTGATATCATTGAAGA 200882

Qy     538 TACTGGTGATGCTTATATCGTTCCTCATGGAGATCATTACCATTACATTCCTAAGAATGA 597
           |||||||||||| ||||||||||||||||  |||| | ||| || |||||||||| |||
Db  200883 TACTGGTGATGCCTATATCGTTCCTCATGGCAATCACTTCCACTATATTCCTAAGAGTGA 200942

Qy     598 GTTATCAGCTAGCGAGTTGGCTGCTGCAGAAGCCTTCCTATCTGGTCGAGGAAATCTGTC 657
           ||||| || || || || |||||||||  ||||||       |||
Db  200943 TTTATCTGCCAGTGAATTAGCTGCTGCCCAAGCCTATTGGAATGG---------------- 200987

Qy     658 AAATTCAAGAACCTATCGCCGACAAAATAGCGATAACACTTCAAGAACAAACTGGGTACC 717
             |  || | | ||    |      ||   | |       |  | | | | ||
Db  200988 -GAAGCAGGGATCTCGTCCTTCTTCAAGTTCTAGTCATAATGCAAATCCAGCTCAACCAA 201046

Qy     718 TTCTGTAAGCAATCCAGGAACTACAAATACTAACACAAGCAACAACAGCAACACTAACAG 777
             ||| ||  | |||  | ||                    |    ||   | |
Db  201047 GATTGTCAGAGAACCACAATCT-------------------GACTGTCACTCCAACTTA 201086

Qy     778 TCAAGCAAGTCAAAGTAATGACATTGATAGTCTCTTGAAACAGCTCTACAAACTGCCTTT 837
           |||   || |||| |  |  |||||   || || ||     |  | ||       || ||
Db  201087 TCATCAAAATCAAGGGGAAAACATTTCAAGCCTTTTACGTGAATTGTATGCTAAACCCTT 201146
```

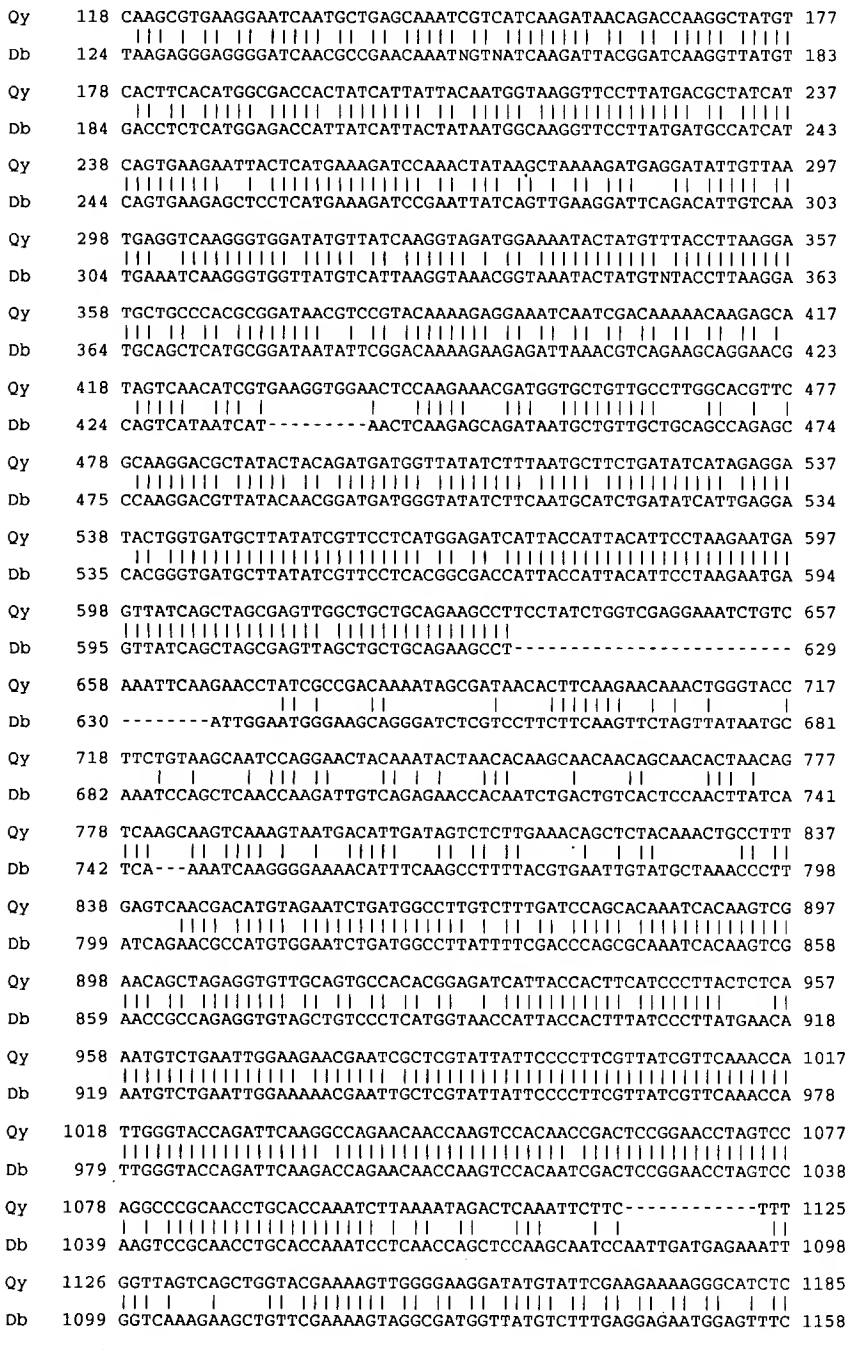
```
Qy     838 GAGTCAACGACATGTAGAATCTGATGGCCTTGTCTTTGATCCAGCACAAATCACAAGTCG 897
Db  201147 ATCAGAACGCCATGTGGAATCTGATGGCCTTATTTTCGACCCAGCGCAAATCACAAGTCG 201206

Qy     898 AACAGCTAGAGGTGTTGCAGTGCCACACGGAGATCATTACCACTTCATCCCTTACTCTCA 957
Db  201207 AACCGCCAGAGGTGTAGCTGTCCCTCATGGTAACCATTACCACTTTATCCCTTATGAACA 201266

Qy     958 AATGTCTGAATTGGAAGAACGAATCGCTCGTATTATTCCCCTTCGTTATCGTTCAAACCA 1017
Db  201267 AATGTCTGAATTGGAAAAACGAATTGCTCGTATTATTCCCCTTCGTTATCGTTCAAACCA 201326

Qy    1018 TTGGGTACCAGATTCAAGGCCAGAACAACCAAGTCCACAACCGACTCCGGAACCTAGTCC 1077
Db  201327 TTGGGTACCAGATTCAAGACCAGAACAACCAAGTCCACAACCGACTCCAGAACCTAGTCC 201386

Qy    1078 AGGCCCGCAACCTGCACCAAATCTTAAAATAGACTCAA-----------ATTCTTCTTT 1125
Db  201387 AAGTCCGCAACCTGCACCAAATCCTCAACCAGCTCCAAGCAATCCAATTGATGAGAAATT 201446

Qy    1126 GGTTAGTCAGCTGGTACGAAAAGTTGGGGAAGGATATGTATTCGAAGAAAAGGGCATCTC 1185
Db  201447 GGTCAAAGAAGCTGTTCGAAAAGTAGGCGATGGTTATGTCTTTGAGGAGAATGGAGTTTC 201506

Qy    1186 TCGTTATGTCTTTGCGAAAGATTTACCATCTGAAACTGTTAAAAATCTTGAAAGCAAGTT 1245
Db  201507 TCGTTATATCCCAGCCAAGGATCTTTCAGCAGAAACAGCAGCAGGCATTGATAGCAAACT 201566

Qy    1246 ATCAAAACAAGAGAGTGTTTCACACACTTTAACTGCTAAAAAAGAAAATGTTGCTCCTCG 1305
Db  201567 GGCCAAGCAGGAAAGTTTATCTCATAAGCTAGGAGCTAAGAAAACTGACCTCCCATCTAG 201626

Qy    1306 TGACCAAGAATTTTATGATAAAGCATATAATCTGTTAACTGAGGCTCATAAAGCCTTGTT 1365
Db  201627 TGATCGAGAATTTTACAATAAGGCTTATGACTTACTAGCAAGAATTCACCAAGATTTACT 201686

Qy    1366 TGNAAATAAGGGTCGTAATTCTGATTTCCAAGCCTTAGACAAATTATTAGAACGCTTGAA 1425
Db  201687 TGATAATAAAGGTCGACAAGTTGATTTTGAGGCTTTGGATAACCTGTTGGAACGACTCAA 201746

Qy    1426 TGATGAATCGACTAATAAAGAAAAATTGGTAGATGATTTATTGGCATTCCTAGCACCAAT 1485
Db  201747 GGATGTCTCAAGTGATAAAGTCAAGTTAGTGGAAGATATTCTTGCCTTCTTAGCTCCGAT 201806

Qy    1486 TACCCATCCAGAGCGACTTGGCAAACCAAATTCTCAAATTGAGTATACTGAAGACGAAGT 1545
Db  201807 TCGTCATCCAGAACGTTTAGGAAAACCAAATGCGCAAATTACCTACACTGATGATGAGAT 201866

Qy    1546 TCGTATTGCTCAATTAGCTGATAAGTATACAACGTCAGATGGTTACATTTTTGATGAACA 1605
Db  201867 TCAAGTAGCCAAGTTGGCAGGCAAGTACACAACAGAAGACGGTTATATCTTTGATCCTCG 201926

Qy    1606 TGATATAATCAGTGATGAAGGAGATGCATATGTAACGCCTCATATGGGCCATAGTCACTG 1665
Db  201927 TGATATAACCAGTGATGAGGGGGATGCCTATGTAACTCCACATATGACCCATAGCCACTG 201986

Qy    1666 GATTGGAAAAGATAGCCTTTCTGATAAGGAAAAAGTTGCAGCTCAAGCCTATACTAAAGA 1725
Db  201987 GATTAAAAAAGATAGTTTGTCTGAAGCTGAAAGAGCGGCAGCCCAGGCTTATGCTAAAGA 202046

Qy    1726 AAAAGGTATCCTACCTCCATCTCCAGACGCAGATGTTAAAGCAAATCCAACTGGAGATAG 1785
Db  202047 GAAAGGTTTGACCCCTCCTTCGACAGACCATCAGGATTCAGGAAATACTGAGGCAAAAGG 202106

Qy    1786 TGCAGCAGCTATTTACAATCGTGTGAAAGGGGAAAAACGAATTCCACTCGTTCGACTTCC 1845
Db  202107 AGCAGAAGCTATCTACAACCGCGTGAAAGCAGCTAAGAAGGTGCCACTTGATCGCATGCC 202166

Qy    1846 ATATATGGTTGAGCATACAGTTGAGGTTAAAAACGGTAATTTGATTATTCCTCATAAGGA 1905
Db  202167 TTACAATCTTCAATATACTGTAGAAGTCAAAAACGGTAGTTTAATCATACCTCATTATGA 202226

Qy    1906 TCATTACCATAATATTAAATTTGCTTGGTTTGATGATCACACATACAAAGCTCCAAATGG 1965
Db  202227 CCATTACCATAACATCAAATTTGAGTGGTTTGACGAAGGCCTTTATGAGGCACCTAAGGG 202286

Qy    1966 CTATACCTTGGAAGATTTGTTTGCGACGATTAAGTACTACGTAGAACACCCTGACGAACG 2025
Db  202287 GTATACTCTTGAGGATCTTTTGGCGACTGTCAAGTACTATGTCGAACATCCAAACGAACG 202346

Qy    2026 TCCACATTCTAATGATGGATGGGGCAATGCCAGTGAGCATGTGTTAGGCAAGAAAGACCA 2085
Db  202347 TCCGCATTCAGATAATGGTTTTGGTAACGCTAGCGACCATGTTCAAAGAAACAAAAATGG 202406

Qy    2086 CAGTGAAGATCCAAATAAGAACTTCAAAGCGGATGAAGAG---------------------- 2125
Db  202407 TCAAGCTGATACCAATCAAACGGAAAAACCAAACGAGGAGAAACCTCAGACAGAAAAACC 202466

Qy    2126 ------------------------------------------------------------ 2125
Db  202467 TGAGGAAGAAACCCCTCGAGAAGAGAAACCGCAAAGCGAGAAACCAGAGTCTCCAAAACC 202526

Qy    2126 ----------CCAGTAGAGGAAACACCTGCTGAGCCAGAAGTCCCTCAAGTAGAGACTGA 2175
Db  202527 AACAGACGAACCAGAAGAAGAATCACCAGAGGAATCAGAAGAACCTCAGGTCGAGACTGA 202586

Qy    2176 AAAAGTAGAAGCCCAACTCAAAGAAGCAGAAGTTTTGCTTGCGAAAGTAACGGATTCTAG 2235
Db  202587 AAAGGTTGAAGAAAAACTGAGAGAGGCTGAAGATTTACTTGGAAAAATCCAGGATCCAAT 202646

Qy    2236 TCTGAAAGCCAATGCAACAGAAACTCTAGCTGGTTTACGAAATAATTTGACTCTTCAAAT 2295
Db  202647 TATCAAGTCCAATGCCAAAGAGACTCTCACAGGATTAAAAAATAATTTACTATTTGGCAC 202706

Qy    2296 TATGGATAACAATAGTATCATGGCAGAAGCAGAAAAATTACTTGCGTTGTTAAAAGGAAG 2355
Db  202707 CCAGGACAACAATACTATTATGGCAGAAGCTGAAAAACTATTGGCTTTATTAAAGGAGAG 202766

Qy    2356 TAA 2358
Db  202767 TAA 202769
```

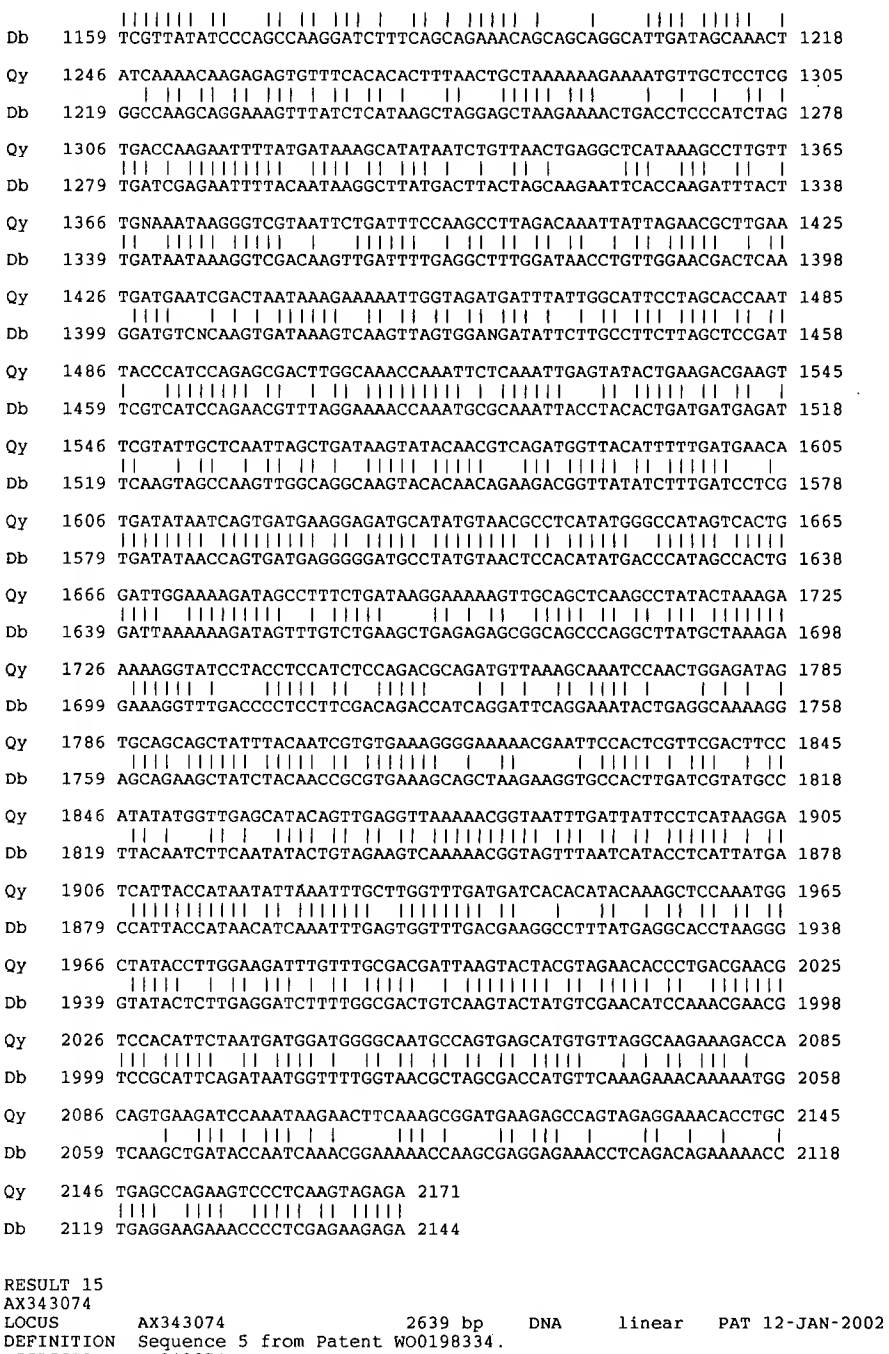
RESULT 14
AR120270

LOCUS	AR120270	2290 bp DNA	linear PAT 16-MAY-2001
DEFINITION	Sequence 65 from patent US 6159469.		
ACCESSION	AR120270		
VERSION	AR120270.1 GI:14103846		
KEYWORDS	.		
SOURCE	Unknown.		
ORGANISM	Unknown.		
	Unclassified.		
REFERENCE	1 (bases 1 to 2290)		
AUTHORS	Choi,G.H., Kunsch,C.A., Barash,S.C., Dillon,P.J., Dougherty,B., Fannon,M.R. and Rosen,C.A.		
TITLE	Streptococcus pneumoniae antigens and vaccines		
JOURNAL	Patent: US 6159469-A 65 12-DEC-2000;		
FEATURES	Location/Qualifiers		
source	1. .2290		
	/organism="unknown"		
BASE COUNT	766 a 474 c 498 g 547 t	5 others	
ORIGIN			

Query Match 41.3%; Score 987.6; DB 6; Length 2290;
Best Local Similarity 67.7%; Pred. No. 1e-200;
Matches 1481; Conservative 0; Mismatches 645; Indels 60; Gaps 5;

```
Qy      1 TTCTTACGAGTTGGGACTGTATCAAGCTAGAACGGTTAAGGAAAA---TAATCGTGTTTC 57
Db      4 TTCCTATGAACTTGGTCGTCACCAAGCTGGTCAGGTTAAGAAAGAGTCTAATCGAGTTTC 63

Qy     58 CTATATAGATGGAAAACAAGCGACGCAAAAAACGGAGAATTTGACTCCTGATGAGGTTAG 117
Db     64 TTATATAGATGGTGATCAGGCTGGTCAAAAGGCAGAAAACTTGACACCAGATGAAGTCAG 123
```





RESULT 15
AX343074
LOCUS AX343074 2639 bp DNA linear PAT 12-JAN-2002
DEFINITION Sequence 5 from Patent WO0198334.
ACCESSION AX343074

```
VERSION     AX343074.1  GI:18152272
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1
  AUTHORS   Hamel,J., Ouellet,C., Charland,N., Martin,D. and Brodeur,B.
  TITLE     Streptococcus antigens
  JOURNAL   Patent: WO 0198334-A 5 27-DEC-2001;
            SHIRE BIOCHEM INC. (CA)
FEATURES             Location/Qualifiers
     source          1. .2639
                     /organism="unidentified"
                     /db_xref="taxon:32644"
                     /note="BVH-11-2"
BASE COUNT      889 a    518 c    567 g    665 t
ORIGIN

  Query Match             41.1%;  Score 980.8;  DB 6;  Length 2639;
  Best Local Similarity   67.8%;  Pred. No. 3e-199;
  Matches 1473;  Conservative    0;  Mismatches  638;  Indels   60;  Gaps    5;

Qy       1 TTCTTACGAGTTGGGACTGTATCAAGCTAGAACGGTTAAGGAAAA---TAATCGTGTTTC 57
Db     173 TTCCTATGAACTTGGTCGTCACCAAGCTGGTCAGGTTAAGAAAGAGTCTAATCGAGTTTC 232

Qy      58 CTATATAGATGGAAAACAAGCGACGCAAAAAACGGAGAATTTGACTCCTGATGAGGTTAG 117
Db     233 TTATATAGATGGTGATCAGGCTGGTCAAAAGGCAGAAAATTTGACACCAGATGAAGTCAG 292

Qy     118 CAAGCGTGAAGGAATCAATGCTGAGCAAATCGTCATCAAGATAACAGACCAAGGCTATGT 177
Db     293 TAAGAGAGAGGGGATCAACGCCGAACAAATTGTTATCAAGATTACGGATCAAGGTTATGT 352

Qy     178 CACTTCACATGGCGACCACTATCATTATTACAATGGTAAGGTTCCTTATGACGCTATCAT 237
Db     353 GACCTCTCATGGAGACCATTATCATTACTATAATGGCAAGGTTCCTTATGATGCCATCAT 412

Qy     238 CAGTGAAGAATTACTCATGAAAGATCCAAACTATAAGCTAAAAGATGAGGATATTGTTAA 297
Db     413 CAGTGAAGAACTTCTCATGAAAGATCCGAATTATCAGTTGAAGGATTCAGACATTGTCAA 472

Qy     298 TGAGGTCAAGGGTGGATATGTTATCAAGGTAGATGGAAAATACTATGTTTACCTTAAGGA 357
Db     473 TGAAATCAAGGGTGGCTATGTGATTAAGGTAGACGGAAAATACTATGTTTACCTTAAAGA 532

Qy     358 TGCTGCCCACGCGGATAACGTCCGTACAAAAGAGGAAATCAATCGACAAAAACAAGAGCA 417
Db     533 TGCGGCCCATGCGGACAATATTCGGACAAAAGAAGAGATTAAACGTCAGAAGCAGGAACA 592

Qy     418 TAGTCAACATCGTGAAGGTGGAACTCCAAGAAACGATGGTGCTGTTGCCTTGGCACGTTC 477
Db     593 CAGTCATAATCAT---------AACTCAAGAGCAGATAATGCTGTTGCTGCAGCCAGAGC 643

Qy     478 GCAAGGACGCTATACTACAGATGATGGTTATATCTTTAATGCTTCTGATATCATAGAGGA 537
Db     644 CCAAGGACGTTATACAACGGATGATGGGTATATCTTCAATGCATCTGATATCATTGAGGA 703

Qy     538 TACTGGTGATGCTTATATCGTTCCTCATGGAGATCATTACCATTACATTCCTAAGAATGA 597
Db     704 CACGGGTGATGCTTATATCGTTCCTCACGGCGACCATTACCATTACATTCCTAAGAATGA 763

Qy     598 GTTATCAGCTAGCGAGTTGGCTGCTGCAGAAGCCTTCCTATCTGGTCGAGGAAATCTGTC 657
Db     764 GTTATCAGCTAGCGAGTTAGCTGCTGCAGAAGCCT---------------------------- 798

Qy     658 AAATTCAAGAACCTATCGCCGACAAAATAGCGATAACACTTCAAGAACAAACTGGGTACC 717
Db     799 --------ATTGGAATGGGAAGCAGGGATCTCGTCCTTCTTCAAGTTCTAGTTATAATGC 850

Qy     718 TTCTGTAAGCAATCCAGGAACTACAAATACTAACACAAGCAACAACAGCAACACTAACAG 777
Db     851 AAATCCAGTTCAACCAAGATTGTCAGAGAACCACAATCTGACTGTCACTCCAACTTATCA 910

Qy     778 TCAAGCAAGTCAAAGTAATGACATTGATAGTCTCTTGAAACAGCTCTACAAACTGCCTTT 837
Db     911 TCA---AAATCAAGGGGAAAACATTTCAAGCCTTTTACGTGAATTGTATGCTAAACCCTT 967

Qy     838 GAGTCAACGACATGTAGAATCTGATGGCCTTGTCTTTGATCCAGCACAAATCACAAGTCG 897
Db     968 ATCAGAACGCCATGTAGAATCTGATGGCCTTATTTTCGACCCAGCGCAAATCACAAGTCG 1027

Qy     898 AACAGCTAGAGGTGTTGCAGTGCCACACGGAGATCATTACCACTTCATCCCTTACTCTCA 957
Db    1028 AACCGCCAGAGGTGTAGCTGTCCCTCATGGTAACCATTACCACTTTATCCCTTATGAACA 1087

Qy     958 AATGTCTGAATTGGAAGAACGAATCGCTCGTATTATTCCCCTTCGTTATCGTTCAAACCA 1017
Db    1088 AATGTCTGAATTGGAAAAACGAATTGCTCGTATTATTCCCCTTCGTTATCGTTCAAACCA 1147

Qy    1018 TTGGGTACCAGATTCAAGGCCAGAACAACCAAGTCCACAACCGACTCCGGAACCTAGTCC 1077
Db    1148 TTGGGTACCAGATTCAAGACCAGAACAACCAAGTCCACAATCGACTCCGGAACCTAGTCC 1207

Qy    1078 AGGCCCGCAACCTGCACCAAATCTTAAAATAGACTCAAATTCTTC------------TTT 1125
Db    1208 AAGTCTGCAACCTGCACCAAATCCTCAACCAGCTCCAAGCAATCCAATTGATGAGAAATT 1267

Qy    1126 GGTTAGTCAGCTGGTACGAAAAGTTGGGGAAGGATATGTATTCGAAGAAAAGGGCATCTC 1185
Db    1268 GGTCAAAGAAGCTGTTCGAAAAGTAGGCGATGGTTATGTCTTTGAGGAGAATGGAGTTTC 1327

Qy    1186 TCGTTATGTCTTTGCGAAAGATTTACCATCTGAAACTGTTAAAAATCTTGAAAGCAAGTT 1245
Db    1328 TCGTTATATCCCAGCCAAGGATCTTTCAGCAGAAACAGCAGCAGGCATTGATAGCAAACT 1387

Qy    1246 ATCAAAACAAGAGAGTGTTTCACACACTTTAACTGCTAAAAAAGAAAATGTTGCTCCTCG 1305
Db    1388 GGCCAAGCAGGAAAGTTTATCTCATAAGCTAGGAGCTAAGAAAACTGACCTCCCATCTAG 1447

Qy    1306 TGACCAAGAATTTTATGATAAAGCATATAATCTGTTAACTGAGGCTCATAAAGCCTTGTT 1365
Db    1448 TGATCGAGAATTTTACAATAAGGCTTATGACTTACTAGCAAGAATTCACCAAGATTTACT 1507

Qy    1366 TGNAAATAAGGGTCGTAATTCTGATTTCCAAGCCTTAGACAAATTATTAGAACGCTTGAA 1425
Db    1508 TGATAATAAAGGTCGACAAGTTGATTTTGAGGTTTTGGATAACCTGTTGGAACGACTCAA 1567

Qy    1426 TGATGAATCGACTAATAAAGAAAAATTGGTAGATGATTTATTGGCATTCCTAGCACCAAT 1485
Db    1568 GGATGTCTCAAGTGATAAAGTCAAGTTAGTGGATGATATTCTTGCCTTCTTAGCTCCGAT 1627

Qy    1486 TACCCATCCAGAGCGACTTGGCAAACCAAATTCTCAAATTGAGTATACTGAAGACGAAGT 1545
Db    1628 TCGTCATCCAGAACGTTTAGGAAAACCAAATGCGCAAATTACCTACACTGATGATGAGAT 1687

Qy    1546 TCGTATTGCTCAATTAGCTGATAAGTATACAACGTCAGATGGTTACATTTTTGATGAACA 1605
Db    1688 TCAAGTAGCCAAGTTGGCAGGCAAGTACACAACAGAAGACGGTTATATCTTTGATCCTCG 1747

Qy    1606 TGATATAATCAGTGATGAAGGAGATGCATATGTAACGCCTCATATGGGCCATAGTCACTG 1665
Db    1748 TGATATAACCAGTGATGAGGGGGATGCCTATGTAACTCCACATATGACCCATAGCCACTG 1807

Qy    1666 GATTGGAAAAGATAGCCTTTCTGATAAGGAAAAAGTTGCAGCTCAAGCCTATACTAAAGA 1725
Db    1808 GATTAAAAAAGATAGTTTGTCTGAAGCTGAGAGAGCGGCAGCCCAGGCTTATGCTAAAGA 1867

Qy    1726 AAAAGGTATCCTACCTCCATCTCCAGACGCAGATGTTAAAGCAAATCCAACTGGAGATAG 1785
Db    1868 GAAAGGTTTGACCCCTCCTTCGACAGACCACCAGGATTCAGGAAATACTGAGGCAAAAGG 1927

Qy    1786 TGCAGCAGCTATTTACAATCGTGTGAAAGGGGAAAAACGAATTCCACTCGTTCGACTTCC 1845
Db    1928 AGCAGAAGCTATCTACAACCGCGTGAAAGCAGCTAAGAAGGTGCCACTTGATCGTATGCC 1987
```

```
Qy   1846 ATATATGGTTGAGCATACAGTTGAGGTTAAAAACGGTAATTTGATTATTCCTCATAAGGA 1905
          || |   || |  |||| || || || |||||||||| ||| || || |||||| | ||
Db   1988 TTACAATCTTCAATATACTGTAGAAGTCAAAAACGGTAGTTTAATCATACCTCATTATGA 2047

Qy   1906 TCATTACCATAATATTAAATTTGCTTGGTTTGATGATCACACATACAAAGCTCCAAATGG 1965
          |||||||||||| || |||||||  |||||||| ||   |   ||  | || || || ||
Db   2048 CCATTACCATAACATCAAATTTGAGTGGTTTGACGAAGGCCTTTATGAGGCACCTAAGGG 2107

Qy   1966 CTATACCTTGGAAGATTTGTTTGCGACGATTAAGTACTACGTAGAACACCCTGACGAACG 2025
          ||||   | || ||| | || |||||  | |||||||| || |||||  ||  |||||||
Db   2108 GTATAGTCTTGAGGATCTTTTGGCGACTGTCAAGTACTATGTCGAACATCCAAACGAACG 2167

Qy   2026 TCCACATTCTAATGATGGATGGGGCAATGCCAGTGAGCATGTGTTAGGCAAGAAAGACCA 2085
          ||| |||||  || |||| |  || || || ||||| |||||       || || |   |
Db   2168 TCCGCATTCAGATAATGGTTTTGGTAACGCTAGTGACCATGTTCGTAAAAATAAGGCAGA 2227

Qy   2086 CAGTGAAGATCCAAATAAGAACTTCAAAGCGGATGAAGAGCCAGTAGAGGAAACACCTGC 2145
          |   ||   |  |  | |  |    || |   |||| ||   |   |||  ||| |   |
Db   2228 CCAAGATAGTAAACCTGATGAAGATAAGGAACATGATGAAGTAAGTGAGCCAACTCACCC 2287

Qy   2146 TGAGCCAGAAG 2156
          |||  | || |
Db   2288 TGAATCTGATG 2298
```

Search completed: November 17, 2002, 03:10:09
Job time : 4780 secs

GenCore version 5.1.3

OM nucleic - nucleic search, using sw model

Run on: November 16, 2002, 23:10:35 ; Search time 340 Seconds
(without alignments)
15823.600 Million cell updates/sec

Title: US-08-961-083-55
Perfect score: 2389
Sequence: 1 TTCTTACGAGTTGGGACTGT.....TAAGTAAGGAAAAAATAAAC 2389

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_101002:*
1: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	2388	100.0	2389	19	AAV27351	Streptococcus pneu
	2	2388	100.0	2389	24	ABQ84819	S. pneumoniae SP03
	3	2388	100.0	2451	21	AAA47604	Recombinant varian
	4	2388	100.0	8195	19	AAV52227	Streptococcus pneu
	5	1374.2	57.5	2523	21	AAA65731	Streptococcus pneu
	6	1374.2	57.5	2647	21	AAA65736	Streptococcus pneu
	7	1374.2	57.5	2647	24	ABK15103	DNA encoding Strep
	8	1011.8	42.4	2478	21	AAA08557	S. pneumoniae 92 k
	9	1003.8	42.0	2481	21	AAA05417	Streptococcus pneu
	10	991	41.5	2531	21	AAA47605	Recombinant varian
	11	990	41.4	2531	21	AAA47602	Recombinant varian
	12	987.6	41.3	2290	19	AAV27356	Streptococcus pneu
	13	987.6	41.3	2290	24	ABQ84824	S. pneumoniae SP04
	14	980.8	41.1	2639	21	AAA65737	Streptococcus pneu
	15	980.8	41.1	2639	24	ABK15104	DNA encoding Strep
	16	869.4	36.4	2163	20	AAX25394	Streptococcus pneu
	17	653.8	27.4	2359	19	AAV52376	Streptococcus pneu
	18	548.4	23.0	3171	21	AAA65739	Streptococcus pneu
	19	385.4	16.1	1342	19	AAV27414	Streptococcus pneu
	20	385.4	16.1	1342	24	ABQ84882	S. pneumoniae SP10
	21	385.4	16.1	1455	21	AAA65733	Streptococcus pneu
	22	385.4	16.1	1455	21	AAA47603	Recombinant varian
	23	385.4	16.1	1455	21	AAA05473	Streptococcus pneu
	24	385.4	16.1	1455	21	AAZ91804	Streptococcus pneu
	25	385.4	16.1	3120	21	AAA65730	Streptococcus pneu
	26	385.4	16.1	5048	21	AAA65735	Streptococcus pneu
	27	385.4	16.1	5048	24	ABK15101	DNA encoding Strep
c	28	385.4	16.1	6867	19	AAV52325	Streptococcus pneu
	29	381.4	16.0	973	19	AAV52488	Streptococcus pneu
	30	347	14.5	2528	21	AAA65738	Streptococcus pneu
	31	347	14.5	2528	24	ABK15105	DNA encoding Strep
	32	247.4	10.4	504	21	AAA08556	S. pneumoniae 20 k
	33	243.4	10.2	1684	19	AAV52391	Streptococcus pneu
	34	214.6	9.0	2475	24	ABN66838	Streptococcus poly
	35	214.6	9.0	2478	22	AAS00036	Streptococcus pyog
	36	211.4	8.8	1146	21	AAA05814	Group B Streptococ
	37	211.4	8.8	2466	24	ABN69535	Streptococcus poly
	38	211.4	8.8	2466	24	ABN70334	Streptococcus poly
	39	211.4	8.8	2469	21	AAA05811	Group B Streptococ
	40	211.4	8.8	2469	22	AAS00038	Streptococcus agal
	41	209.8	8.8	2469	21	AAA65740	Streptococcus pneu
	42	209.8	8.8	5215	20	AAX91105	Group B Streptococ
	43	206.6	8.6	2472	21	AAA65741	Streptococcus pneu
	44	201.6	8.4	1398	24	ABN66839	Streptococcus poly
	45	201.2	8.4	819	24	ABK15106	DNA encoding Strep

ALIGNMENTS

RESULT 1
AAV27351
ID AAV27351 standard; DNA; 2389 BP.
XX
AC AAV27351;
XX
DT 02-OCT-1998 (first entry)
XX
DE Streptococcus pneumoniae SP0036 nucleotide.
XX
KW Streptococcus pneumoniae; antigen; vaccine; infection; diagnosis;
KW detection; pneumonia; otitis media; meningitis; ss.
XX
OS Streptococcus pneumoniae.
XX
FH Key Location/Qualifiers
FT CDS 2..2389
FT /*tag= a
FT /product= "SP0036"
FT /transl_except= (pos:1367..1369,aa:Xaa)
FT /note= "no stop codon given; Xaa is unspecified"
XX
PN WO9818930-A2.
XX
PD 07-MAY-1998.
XX
PF 30-OCT-1997; 97WO-US19422.
XX
PR 31-OCT-1996; 96US-0029960.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX

PI Choi GH, Hromockyj A, Johnson LS, Kunsch CA;
XX
DR WPI; 1998-272224/24.
DR P-PSDB; AAW55090.
XX
PT Nucleic acid encoding antigenic peptide(s) from Streptococcus
PT pneumoniae - or their epitope-containing fragments, useful in
PT protective or therapeutic vaccines, and for diagnosis
XX
PS Claim 1; Page 59; 118pp; English.
XX
CC The present sequence encodes a protein from Streptococcus pneumoniae.
CC The nucleic acid sequence encoding the Streptococcus pneumoniae protein
CC can be useful in vaccines for inducing protective antibodies against
CC Streptococcus pneumoniae, for treatment or prevention of infection e.g.
CC pneumonia, otitis media or meningitis. Probes based on the nucleic acid
CC are used to detect Streptococcus infection (by usual hybridisation or
CC amplification methods), also for isolating Streptococcus genes or their
CC allelic variants. The protein can be used similarly to detect specific
CC antibodies in standard immunoassays, especially for diagnosing or
CC monitoring infections. Antibodies which bind the protein are used to
CC detect corresponding antigens, to purify the protein and for passive
CC immunisation (optionally coupled to a toxin). Vaccines are administered,
CC e.g. by injection, orally or through the skin, typically at 0.01-1000
CC (especially 10-300) mu g/ml per dose.
XX
SQ Sequence 2389 BP; 830 A; 461 C; 486 G; 611 T; 1 other;

Query Match 100.0%; Score 2388; DB 19; Length 2389;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2389; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 TTCTTACGAGTTGGGACTGTATCAAGCTAGAACGGTTAAGGAAAATAATCGTGTTTCCTA 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 TTCTTACGAGTTGGGACTGTATCAAGCTAGAACGGTTAAGGAAAATAATCGTGTTTCCTA 60

Qy      61 TATAGATGGAAAACAAGCGACGCAAAAAACGGAGAATTTGACTCCTGATGAGGTTAGCAA 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 TATAGATGGAAAACAAGCGACGCAAAAAACGGAGAATTTGACTCCTGATGAGGTTAGCAA 120

Qy     121 GCGTGAAGGAATCAATGCTGAGCAAATCGTCATCAAGATAACAGACCAAGGCTATGTCAC 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 GCGTGAAGGAATCAATGCTGAGCAAATCGTCATCAAGATAACAGACCAAGGCTATGTCAC 180

Qy     181 TTCACATGGCGACCACTATCATTATTACAATGGTAAGGTTCCTTATGACGCTATCATCAG 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 TTCACATGGCGACCACTATCATTATTACAATGGTAAGGTTCCTTATGACGCTATCATCAG 240

Qy     241 TGAAGAATTACTCATGAAAGATCCAAACTATAAGCTAAAAGATGAGGATATTGTTAATGA 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 TGAAGAATTACTCATGAAAGATCCAAACTATAAGCTAAAAGATGAGGATATTGTTAATGA 300

Qy     301 GGTCAAGGGTGGATATGTTATCAAGGTAGATGGAAAATACTATGTTTACCTTAAGGATGC 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 GGTCAAGGGTGGATATGTTATCAAGGTAGATGGAAAATACTATGTTTACCTTAAGGATGC 360

Qy     361 TGCCCACGCGGATAACGTCCGTACAAAAGAGGAAATCAATCGACAAAAACAAGAGCATAG 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 TGCCCACGCGGATAACGTCCGTACAAAAGAGGAAATCAATCGACAAAAACAAGAGCATAG 420

Qy     421 TCAACATCGTGAAGGTGGAACTCCAAGAAACGATGGTGCTGTTGCCTTGGCACGTTCGCA 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 TCAACATCGTGAAGGTGGAACTCCAAGAAACGATGGTGCTGTTGCCTTGGCACGTTCGCA 480

Qy     481 AGGACGCTATACTACAGATGATGGTTATATCTTTAATGCTTCTGATATCATAGAGGATAC 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 AGGACGCTATACTACAGATGATGGTTATATCTTTAATGCTTCTGATATCATAGAGGATAC 540

Qy     541 TGGTGATGCTTATATCGTTCCTCATGGAGATCATTACCATTACATTCCTAAGAATGAGTT 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     541 TGGTGATGCTTATATCGTTCCTCATGGAGATCATTACCATTACATTCCTAAGAATGAGTT 600

Qy     601 ATCAGCTAGCGAGTTGGCTGCTGCAGAAGCCTTCCTATCTGGTCGAGGAAATCTGTCAAA 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     601 ATCAGCTAGCGAGTTGGCTGCTGCAGAAGCCTTCCTATCTGGTCGAGGAAATCTGTCAAA 660

Qy     661 TTCAAGAACCTATCGCCGACAAAATAGCGATAACACTTCAAGAACAAACTGGGTACCTTC 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     661 TTCAAGAACCTATCGCCGACAAAATAGCGATAACACTTCAAGAACAAACTGGGTACCTTC 720

Qy     721 TGTAAGCAATCCAGGAACTACAAATACTAACACAAGCAACAACAGCAACACTAACAGTCA 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     721 TGTAAGCAATCCAGGAACTACAAATACTAACACAAGCAACAACAGCAACACTAACAGTCA 780

Qy     781 AGCAAGTCAAAGTAATGACATTGATAGTCTCTTGAAACAGCTCTACAAACTGCCTTTGAG 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     781 AGCAAGTCAAAGTAATGACATTGATAGTCTCTTGAAACAGCTCTACAAACTGCCTTTGAG 840

Qy     841 TCAACGACATGTAGAATCTGATGGCCTTGTCTTTGATCCAGCACAAATCACAAGTCGAAC 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     841 TCAACGACATGTAGAATCTGATGGCCTTGTCTTTGATCCAGCACAAATCACAAGTCGAAC 900

Qy     901 AGCTAGAGGTGTTGCAGTGCCACACGGAGATCATTACCACTTCATCCCTTACTCTCAAAT 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     901 AGCTAGAGGTGTTGCAGTGCCACACGGAGATCATTACCACTTCATCCCTTACTCTCAAAT 960

Qy     961 GTCTGAATTGGAAGAACGAATCGCTCGTATTATTCCCCTTCGTTATCGTTCAAACCATTG 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     961 GTCTGAATTGGAAGAACGAATCGCTCGTATTATTCCCCTTCGTTATCGTTCAAACCATTG 1020

Qy    1021 GGTACCAGATTCAAGGCCAGAACAACCAAGTCCACAACCGACTCCGGAACCTAGTCCAGG 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1021 GGTACCAGATTCAAGGCCAGAACAACCAAGTCCACAACCGACTCCGGAACCTAGTCCAGG 1080

Qy    1081 CCCGCAACCTGCACCAAATCTTAAAATAGACTCAAATTCTTCTTTGGTTAGTCAGCTGGT 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1081 CCCGCAACCTGCACCAAATCTTAAAATAGACTCAAATTCTTCTTTGGTTAGTCAGCTGGT 1140

Qy    1141 ACGAAAAGTTGGGGAAGGATATGTATTCGAAGAAAAGGGCATCTCTCGTTATGTCTTTGC 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1141 ACGAAAAGTTGGGGAAGGATATGTATTCGAAGAAAAGGGCATCTCTCGTTATGTCTTTGC 1200

Qy    1201 GAAAGATTTACCATCTGAAACTGTTAAAAATCTTGAAAGCAAGTTATCAAAACAAGAGAG 1260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1201 GAAAGATTTACCATCTGAAACTGTTAAAAATCTTGAAAGCAAGTTATCAAAACAAGAGAG 1260

Qy    1261 TGTTTCACACACTTTAACTGCTAAAAAAGAAAATGTTGCTCCTCGTGACCAAGAATTTTA 1320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1261 TGTTTCACACACTTTAACTGCTAAAAAAGAAAATGTTGCTCCTCGTGACCAAGAATTTTA 1320

Qy    1321 TGATAAAGCATATAATCTGTTAACTGAGGCTCATAAAGCCTTGTTTGNAAATAAGGGTCG 1380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1321 TGATAAAGCATATAATCTGTTAACTGAGGCTCATAAAGCCTTGTTTGNAAATAAGGGTCG 1380

Qy    1381 TAATTCTGATTTCCAAGCCTTAGACAAATTATTAGAACGCTTGAATGATGAATCGACTAA 1440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1381 TAATTCTGATTTCCAAGCCTTAGACAAATTATTAGAACGCTTGAATGATGAATCGACTAA 1440

Qy    1441 TAAAGAAAAATTGGTAGATGATTTATTGGCATTCCTAGCACCAATTACCCATCCAGAGCG 1500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1441 TAAAGAAAAATTGGTAGATGATTTATTGGCATTCCTAGCACCAATTACCCATCCAGAGCG 1500

Qy    1501 ACTTGGCAAACCAAATTCTCAAATTGAGTATACTGAAGACGAAGTTCGTATTGCTCAATT 1560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1501 ACTTGGCAAACCAAATTCTCAAATTGAGTATACTGAAGACGAAGTTCGTATTGCTCAATT 1560

Qy    1561 AGCTGATAAGTATACAACGTCAGATGGTTACATTTTTGATGAACATGATATAATCAGTGA 1620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1561 AGCTGATAAGTATACAACGTCAGATGGTTACATTTTTGATGAACATGATATAATCAGTGA 1620

Qy    1621 TGAAGGAGATGCATATGTAACGCCTCATATGGGCCATAGTCACTGGATTGGAAAAGATAG 1680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1621 TGAAGGAGATGCATATGTAACGCCTCATATGGGCCATAGTCACTGGATTGGAAAAGATAG 1680

Qy    1681 CCTTTCTGATAAGGAAAAAGTTGCAGCTCAAGCCTATACTAAAGAAAAAGGTATCCTACC 1740
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db   1681 CCTTTCTGATAAGGAAAAAGTTGCAGCTCAAGCCTATACTAAAGAAAAAGGTATCCTACC 1740

Qy   1741 TCCATCTCCAGACGCAGATGTTAAAGCAAATCCAACTGGAGATAGTGCAGCAGCTATTTA 1800
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1741 TCCATCTCCAGACGCAGATGTTAAAGCAAATCCAACTGGAGATAGTGCAGCAGCTATTTA 1800

Qy   1801 CAATCGTGTGAAAGGGGAAAAACGAATTCCACTCGTTCGACTTCCATATATGGTTGAGCA 1860
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1801 CAATCGTGTGAAAGGGGAAAAACGAATTCCACTCGTTCGACTTCCATATATGGTTGAGCA 1860

Qy   1861 TACAGTTGAGGTTAAAAACGGTAATTTGATTATTCCTCATAAGGATCATTACCATAATAT 1920
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1861 TACAGTTGAGGTTAAAAACGGTAATTTGATTATTCCTCATAAGGATCATTACCATAATAT 1920

Qy   1921 TAAATTTGCTTGGTTTGATGATCACACATACAAAGCTCCAAATGGCTATACCTTGGAAGA 1980
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1921 TAAATTTGCTTGGTTTGATGATCACACATACAAAGCTCCAAATGGCTATACCTTGGAAGA 1980

Qy   1981 TTTGTTTGCGACGATTAAGTACTACGTAGAACACCCTGACGAACGTCCACATTCTAATGA 2040
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1981 TTTGTTTGCGACGATTAAGTACTACGTAGAACACCCTGACGAACGTCCACATTCTAATGA 2040

Qy   2041 TGGATGGGGCAATGCCAGTGAGCATGTGTTAGGCAAGAAAGACCACAGTGAAGATCCAAA 2100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2041 TGGATGGGGCAATGCCAGTGAGCATGTGTTAGGCAAGAAAGACCACAGTGAAGATCCAAA 2100

Qy   2101 TAAGAACTTCAAAGCGGATGAAGAGCCAGTAGAGGAAACACCTGCTGAGCCAGAAGTCCC 2160
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2101 TAAGAACTTCAAAGCGGATGAAGAGCCAGTAGAGGAAACACCTGCTGAGCCAGAAGTCCC 2160

Qy   2161 TCAAGTAGAGACTGAAAAAGTAGAAGCCCAACTCAAAGAAGCAGAAGTTTTGCTTGCGAA 2220
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2161 TCAAGTAGAGACTGAAAAAGTAGAAGCCCAACTCAAAGAAGCAGAAGTTTTGCTTGCGAA 2220

Qy   2221 AGTAACGGATTCTAGTCTGAAAGCCAATGCAACAGAAACTCTAGCTGGTTTACGAAATAA 2280
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2221 AGTAACGGATTCTAGTCTGAAAGCCAATGCAACAGAAACTCTAGCTGGTTTACGAAATAA 2280

Qy   2281 TTTGACTCTTCAAATTATGGATAACAATAGTATCATGGCAGAAGCAGAAAAATTACTTGC 2340
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2281 TTTGACTCTTCAAATTATGGATAACAATAGTATCATGGCAGAAGCAGAAAAATTACTTGC 2340

Qy   2341 GTTGTTAAAAGGAAGTAATCCTTCATCTGTAAGTAAGGAAAAAATAAAC 2389
          |||||||||||||||||||||||||||||||||||||||||||||||||
Db   2341 GTTGTTAAAAGGAAGTAATCCTTCATCTGTAAGTAAGGAAAAAATAAAC 2389
```

RESULT 2
ABQ84819
ID ABQ84819 standard; DNA; 2389 BP.
XX
AC ABQ84819;
XX
DT 04-SEP-2002 (first entry)
XX
DE S. pneumoniae SP036 nucleotide sequence SEQ ID NO:55.
XX
KW Streptococcus pneumoniae; epitope; vaccine; antigenic protein;
KW antibacterial; Streptococcal infection; detection; gene; ds.
XX
OS Streptococcus pneumoniae.
XX
PN US2002061545-A1.
XX
PD 23-MAY-2002.
XX
PF 22-JAN-2001; 2001US-0765272.
XX
PR 30-OCT-1997; 97US-0961083.
XX
PA (CHOI/) CHOI G H.
PA (KUNS/) KUNSCH C A.
PA (BARA/) BARASH S C.
PA (DILL/) DILLON P J.
PA (DOUG/) DOUGHERTY B.
PA (FANN/) FANNON M R.
PA (ROSE/) ROSEN C A.
XX
PI Choi GH, Kunsch CA, Barash SC, Dillon PJ, Dougherty B, Fannon MR;
PI Rosen CA;
XX
DR WPI; 2002-479261/51.
DR P-PSDB; ABP54584.
XX
PT New Streptococcus pneumoniae antigens, useful for detecting
PT Streptococcus and for preventing or attenuating disease caused by
PT Streptococcus infection -
XX
PS Claim 1; Page 27; 70pp; English.
XX
CC ABQ84792 to ABQ84904 represents nucleic acids which encode the
CC Streptococcus pneumoniae antigens given in ABP54557 to ABP54669.
CC The S. pneumoniae antigens have antibacterial activity and can be
CC used in vaccines. The S. pneumoniae antigens can also be used to
CC prevent or attenuate a Streptococcal infection in an animal. The
CC polynucleotides encoding the S. pneumoniae antigens can be used to
CC detect Streptococcus nucleic acids. ABQ84905 to ABQ85130 represent
CC primers used in the cloning of S. pneumoniae ORFs (open reading frames)
CC which are used in an example from the present invention.
XX
SQ Sequence 2389 BP; 830 A; 461 C; 486 G; 611 T; 1 other;

Query Match 100.0%; Score 2388; DB 24; Length 2389;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2389; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      1 TTCTTACGAGTTGGGACTGTATCAAGCTAGAACGGTTAAGGAAAATAATCGTGTTTCCTA 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 TTCTTACGAGTTGGGACTGTATCAAGCTAGAACGGTTAAGGAAAATAATCGTGTTTCCTA 60

Qy     61 TATAGATGGAAAACAAGCGACGCAAAAAACGGAGAATTTGACTCCTGATGAGGTTAGCAA 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 TATAGATGGAAAACAAGCGACGCAAAAAACGGAGAATTTGACTCCTGATGAGGTTAGCAA 120

Qy    121 GCGTGAAGGAATCAATGCTGAGCAAATCGTCATCAAGATAACAGACCAAGGCTATGTCAC 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 GCGTGAAGGAATCAATGCTGAGCAAATCGTCATCAAGATAACAGACCAAGGCTATGTCAC 180

Qy    181 TTCACATGGCGACCACTATCATTATTACAATGGTAAGGTTCCTTATGACGCTATCATCAG 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 TTCACATGGCGACCACTATCATTATTACAATGGTAAGGTTCCTTATGACGCTATCATCAG 240

Qy    241 TGAAGAATTACTCATGAAAGATCCAAACTATAAGCTAAAAGATGAGGATATTGTTAATGA 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 TGAAGAATTACTCATGAAAGATCCAAACTATAAGCTAAAAGATGAGGATATTGTTAATGA 300

Qy    301 GGTCAAGGGTGGATATGTTATCAAGGTAGATGGAAAATACTATGTTTACCTTAAGGATGC 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 GGTCAAGGGTGGATATGTTATCAAGGTAGATGGAAAATACTATGTTTACCTTAAGGATGC 360

Qy    361 TGCCCACGCGGATAACGTCCGTACAAAAGAGGAAATCAATCGACAAAAACAAGAGCATAG 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 TGCCCACGCGGATAACGTCCGTACAAAAGAGGAAATCAATCGACAAAAACAAGAGCATAG 420

Qy    421 TCAACATCGTGAAGGTGGAACTCCAAGAAACGATGGTGCTGTTGCCTTGGCACGTTCGCA 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 TCAACATCGTGAAGGTGGAACTCCAAGAAACGATGGTGCTGTTGCCTTGGCACGTTCGCA 480

Qy    481 AGGACGCTATACTACAGATGATGGTTATATCTTTAATGCTTCTGATATCATAGAGGATAC 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 AGGACGCTATACTACAGATGATGGTTATATCTTTAATGCTTCTGATATCATAGAGGATAC 540

Qy    541 TGGTGATGCTTATATCGTTCCTCATGGAGATCATTACCATTACATTCCTAAGAATGAGTT 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 TGGTGATGCTTATATCGTTCCTCATGGAGATCATTACCATTACATTCCTAAGAATGAGTT 600
```

```
Qy    601 ATCAGCTAGCGAGTTGGCTGCTGCAGAAGCCTTCCTATCTGGTCGAGGAAATCTGTCAAA 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 ATCAGCTAGCGAGTTGGCTGCTGCAGAAGCCTTCCTATCTGGTCGAGGAAATCTGTCAAA 660

Qy    661 TTCAAGAACCTATCGCCGACAAAATAGCGATAACACTTCAAGAACAAACTGGGTACCTTC 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661 TTCAAGAACCTATCGCCGACAAAATAGCGATAACACTTCAAGAACAAACTGGGTACCTTC 720

Qy    721 TGTAAGCAATCCAGGAACTACAAATACTAACACAAGCAACAACAGCAACACTAACAGTCA 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    721 TGTAAGCAATCCAGGAACTACAAATACTAACACAAGCAACAACAGCAACACTAACAGTCA 780

Qy    781 AGCAAGTCAAAGTAATGACATTGATAGTCTCTTGAAACAGCTCTACAAACTGCCTTTGAG 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    781 AGCAAGTCAAAGTAATGACATTGATAGTCTCTTGAAACAGCTCTACAAACTGCCTTTGAG 840

Qy    841 TCAACGACATGTAGAATCTGATGGCCTTGTCTTTGATCCAGCACAAATCACAAGTCGAAC 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    841 TCAACGACATGTAGAATCTGATGGCCTTGTCTTTGATCCAGCACAAATCACAAGTCGAAC 900

Qy    901 AGCTAGAGGTGTTGCAGTGCCACACGGAGATCATTACCACTTCATCCCTTACTCTCAAAT 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    901 AGCTAGAGGTGTTGCAGTGCCACACGGAGATCATTACCACTTCATCCCTTACTCTCAAAT 960

Qy    961 GTCTGAATTGGAAGAACGAATCGCTCGTATTATTCCCCTTCGTTATCGTTCAAACCATTG 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    961 GTCTGAATTGGAAGAACGAATCGCTCGTATTATTCCCCTTCGTTATCGTTCAAACCATTG 1020

Qy   1021 GGTACCAGATTCAAGGCCAGAACAACCAAGTCCACAACCGACTCCGGAACCTAGTCCAGG 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1021 GGTACCAGATTCAAGGCCAGAACAACCAAGTCCACAACCGACTCCGGAACCTAGTCCAGG 1080

Qy   1081 CCCGCAACCTGCACCAAATCTTAAAATAGACTCAAATTCTTCTTTGGTTAGTCAGCTGGT 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1081 CCCGCAACCTGCACCAAATCTTAAAATAGACTCAAATTCTTCTTTGGTTAGTCAGCTGGT 1140

Qy   1141 ACGAAAAGTTGGGGAAGGATATGTATTCGAAGAAAAGGGCATCTCTCGTTATGTCTTTGC 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1141 ACGAAAAGTTGGGGAAGGATATGTATTCGAAGAAAAGGGCATCTCTCGTTATGTCTTTGC 1200

Qy   1201 GAAAGATTTACCATCTGAAACTGTTAAAAATCTTGAAAGCAAGTTATCAAAACAAGAGAG 1260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1201 GAAAGATTTACCATCTGAAACTGTTAAAAATCTTGAAAGCAAGTTATCAAAACAAGAGAG 1260

Qy   1261 TGTTTCACACACTTTAACTGCTAAAAAAGAAAATGTTGCTCCTCGTGACCAAGAATTTTA 1320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1261 TGTTTCACACACTTTAACTGCTAAAAAAGAAAATGTTGCTCCTCGTGACCAAGAATTTTA 1320

Qy   1321 TGATAAAGCATATAATCTGTTAACTGAGGCTCATAAAGCCTTGTTTGNAAATAAGGGTCG 1380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1321 TGATAAAGCATATAATCTGTTAACTGAGGCTCATAAAGCCTTGTTTGNAAATAAGGGTCG 1380

Qy   1381 TAATTCTGATTTCCAAGCCTTAGACAAATTATTAGAACGCTTGAATGATGAATCGACTAA 1440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1381 TAATTCTGATTTCCAAGCCTTAGACAAATTATTAGAACGCTTGAATGATGAATCGACTAA 1440

Qy   1441 TAAAGAAAAATTGGTAGATGATTTATTGGCATTCCTAGCACCAATTACCCATCCAGAGCG 1500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1441 TAAAGAAAAATTGGTAGATGATTTATTGGCATTCCTAGCACCAATTACCCATCCAGAGCG 1500

Qy   1501 ACTTGGCAAACCAAATTCTCAAATTGAGTATACTGAAGACGAAGTTCGTATTGCTCAATT 1560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1501 ACTTGGCAAACCAAATTCTCAAATTGAGTATACTGAAGACGAAGTTCGTATTGCTCAATT 1560

Qy   1561 AGCTGATAAGTATACAACGTCAGATGGTTACATTTTTGATGAACATGATATAATCAGTGA 1620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1561 AGCTGATAAGTATACAACGTCAGATGGTTACATTTTTGATGAACATGATATAATCAGTGA 1620

Qy   1621 TGAAGGAGATGCATATGTAACGCCTCATATGGGCCATAGTCACTGGATTGGAAAAGATAG 1680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1621 TGAAGGAGATGCATATGTAACGCCTCATATGGGCCATAGTCACTGGATTGGAAAAGATAG 1680

Qy   1681 CCTTTCTGATAAGGAAAAAGTTGCAGCTCAAGCCTATACTAAAGAAAAAGGTATCCTACC 1740
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1681 CCTTTCTGATAAGGAAAAAGTTGCAGCTCAAGCCTATACTAAAGAAAAAGGTATCCTACC 1740

Qy   1741 TCCATCTCCAGACGCAGATGTTAAAGCAAATCCAACTGGAGATAGTGCAGCAGCTATTTA 1800
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1741 TCCATCTCCAGACGCAGATGTTAAAGCAAATCCAACTGGAGATAGTGCAGCAGCTATTTA 1800

Qy   1801 CAATCGTGTGAAAGGGGAAAAACGAATTCCACTCGTTCGACTTCCATATATGGTTGAGCA 1860
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1801 CAATCGTGTGAAAGGGGAAAAACGAATTCCACTCGTTCGACTTCCATATATGGTTGAGCA 1860

Qy   1861 TACAGTTGAGGTTAAAAACGGTAATTTGATTATTCCTCATAAGGATCATTACCATAATAT 1920
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1861 TACAGTTGAGGTTAAAAACGGTAATTTGATTATTCCTCATAAGGATCATTACCATAATAT 1920

Qy   1921 TAAATTTGCTTGGTTTGATGATCACACATACAAAGCTCCAAATGGCTATACCTTGGAAGA 1980
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1921 TAAATTTGCTTGGTTTGATGATCACACATACAAAGCTCCAAATGGCTATACCTTGGAAGA 1980

Qy   1981 TTTGTTTGCGACGATTAAGTACTACGTAGAACACCCTGACGAACGTCCACATTCTAATGA 2040
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1981 TTTGTTTGCGACGATTAAGTACTACGTAGAACACCCTGACGAACGTCCACATTCTAATGA 2040

Qy   2041 TGGATGGGGCAATGCCAGTGAGCATGTGTTAGGCAAGAAAGACCACAGTGAAGATCCAAA 2100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2041 TGGATGGGGCAATGCCAGTGAGCATGTGTTAGGCAAGAAAGACCACAGTGAAGATCCAAA 2100

Qy   2101 TAAGAACTTCAAAGCGGATGAAGAGCCAGTAGAGGAAACACCTGCTGAGCCAGAAGTCCC 2160
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2101 TAAGAACTTCAAAGCGGATGAAGAGCCAGTAGAGGAAACACCTGCTGAGCCAGAAGTCCC 2160

Qy   2161 TCAAGTAGAGACTGAAAAAGTAGAAGCCCAACTCAAAGAAGCAGAAGTTTTGCTTGCGAA 2220
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2161 TCAAGTAGAGACTGAAAAAGTAGAAGCCCAACTCAAAGAAGCAGAAGTTTTGCTTGCGAA 2220

Qy   2221 AGTAACGGATTCTAGTCTGAAAGCCAATGCAACAGAAACTCTAGCTGGTTTACGAAATAA 2280
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2221 AGTAACGGATTCTAGTCTGAAAGCCAATGCAACAGAAACTCTAGCTGGTTTACGAAATAA 2280

Qy   2281 TTTGACTCTTCAAATTATGGATAACAATAGTATCATGGCAGAAGCAGAAAAATTACTTGC 2340
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2281 TTTGACTCTTCAAATTATGGATAACAATAGTATCATGGCAGAAGCAGAAAAATTACTTGC 2340

Qy   2341 GTTGTTAAAAGGAAGTAATCCTTCATCTGTAAGTAAGGAAAAAATAAAC 2389
          |||||||||||||||||||||||||||||||||||||||||||||||||
Db   2341 GTTGTTAAAAGGAAGTAATCCTTCATCTGTAAGTAAGGAAAAAATAAAC 2389
```

RESULT 3
AAA47604
ID AAA47604 standard; DNA; 2451 BP.
XX
AC AAA47604;
XX
DT 20-OCT-2000 (first entry)
XX
DE Recombinant variant of Sp36 gene (Sp36A) of S. pneumoniae.
XX
KW Streptococcus pneumoniae; infection; vaccine; coiled coil region;
KW histidine triad residue; Sp36; antibody; otitis media;
KW nasopharyngeal infection; bronchial infection; bronchitis; sepsis;
KW meningitis; lobar pneumonia; ds.
XX
OS Streptococcus pneumoniae.
XX
FH Key Location/Qualifiers
FT CDS 1..2451
FT /*tag= a
FT /product= Sp36A polypeptide
XX
PN WO200037105-A2.
XX
PD 29-JUN-2000.

XX
PF 21-DEC-1999; 99WO-US30390.
XX
PR 21-DEC-1998; 98US-0113048.
XX
PA (MEDI-) MEDIMMUNE INC.
XX
PI Johnson LS, Koenig S, Adamou JE;
XX
DR WPI; 2000-452129/39.
DR P-PSDB; AAB01468.
XX
PT Vaccine useful for prophylaxis and treatment of pnemococcal infections
PT such as otitis media, nasopharyngeal and bronchial infections,
PT comprises Streptococcus pneumoniae proteins
XX
PS Disclosure; Page 64-65; 70pp; English.
XX
CC Although a number of proteins have been suggested as being involved
CC in the pathogenicity of Streptococcus pneumoniae, there still remains
CC a need to identify polypeptides having epitopes in common from
CC various strains of S. pneumoniae in order to utilise such
CC polypeptides in vaccines to protect against a wide variety of
CC S. pneumoniae. New vaccine compositions are described which comprise a
CC Streptococcus pneumoniae polypeptide (or fragments) of 80 - 680 amino
CC acids in length that comprise at least one histidine triad residue
CC (HxxHxH) or a coiled-coil region, or an antibody directed against
CC these features. The vaccine is useful in protecting against infection
CC by Streptococcus pneumoniae. The vaccine composition comprising
CC antibodies to is useful for passive immunization for treating
CC Pneumococcal infections which includes otitis media, nasopharyngeal
CC and bronchial infections.
XX
SQ Sequence 2451 BP; 849 A; 467 C; 499 G; 635 T; 1 other;

Query Match 100.0%; Score 2388; DB 21; Length 2451;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2389; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 TTCTTACGAGTTGGGACTGTATCAAGCTAGAACGGTTAAGGAAAATAATCGTGTTTCCTA 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      60 TTCTTACGAGTTGGGACTGTATCAAGCTAGAACGGTTAAGGAAAATAATCGTGTTTCCTA 119

Qy      61 TATAGATGGAAAACAAGCGACGCAAAAAACGGAGAATTTGACTCCTGATGAGGTTAGCAA 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     120 TATAGATGGAAAACAAGCGACGCAAAAAACGGAGAATTTGACTCCTGATGAGGTTAGCAA 179

Qy     121 GCGTGAAGGAATCAATGCTGAGCAAATCGTCATCAAGATAACAGACCAAGGCTATGTCAC 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     180 GCGTGAAGGAATCAATGCTGAGCAAATCGTCATCAAGATAACAGACCAAGGCTATGTCAC 239

Qy     181 TTCACATGGCGACCACTATCATTATTACAATGGTAAGGTTCCTTATGACGCTATCATCAG 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     240 TTCACATGGCGACCACTATCATTATTACAATGGTAAGGTTCCTTATGACGCTATCATCAG 299

Qy     241 TGAAGAATTACTCATGAAAGATCCAAACTATAAGCTAAAAGATGAGGATATTGTTAATGA 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     300 TGAAGAATTACTCATGAAAGATCCAAACTATAAGCTAAAAGATGAGGATATTGTTAATGA 359

Qy     301 GGTCAAGGGTGGATATGTTATCAAGGTAGATGGAAAATACTATGTTTACCTTAAGGATGC 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     360 GGTCAAGGGTGGATATGTTATCAAGGTAGATGGAAAATACTATGTTTACCTTAAGGATGC 419

Qy     361 TGCCCACGCGGATAACGTCCGTACAAAAGAGGAAATCAATCGACAAAAACAAGAGCATAG 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     420 TGCCCACGCGGATAACGTCCGTACAAAAGAGGAAATCAATCGACAAAAACAAGAGCATAG 479

Qy     421 TCAACATCGTGAAGGTGGAACTCCAAGAAACGATGGTGCTGTTGCCTTGGCACGTTCGCA 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     480 TCAACATCGTGAAGGTGGAACTCCAAGAAACGATGGTGCTGTTGCCTTGGCACGTTCGCA 539

Qy     481 AGGACGCTATACTACAGATGATGGTTATATCTTTAATGCTTCTGATATCATAGAGGATAC 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     540 AGGACGCTATACTACAGATGATGGTTATATCTTTAATGCTTCTGATATCATAGAGGATAC 599

Qy     541 TGGTGATGCTTATATCGTTCCTCATGGAGATCATTACCATTACATTCCTAAGAATGAGTT 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     600 TGGTGATGCTTATATCGTTCCTCATGGAGATCATTACCATTACATTCCTAAGAATGAGTT 659

Qy     601 ATCAGCTAGCGAGTTGGCTGCTGCAGAAGCCTTCCTATCTGGTCGAGGAAATCTGTCAAA 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     660 ATCAGCTAGCGAGTTGGCTGCTGCAGAAGCCTTCCTATCTGGTCGAGGAAATCTGTCAAA 719

Qy     661 TTCAAGAACCTATCGCCGACAAAATAGCGATAACACTTCAAGAACAAACTGGGTACCTTC 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     720 TTCAAGAACCTATCGCCGACAAAATAGCGATAACACTTCAAGAACAAACTGGGTACCTTC 779

Qy     721 TGTAAGCAATCCAGGAACTACAAATACTAACACAAGCAACAACAGCAACACTAACAGTCA 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     780 TGTAAGCAATCCAGGAACTACAAATACTAACACAAGCAACAACAGCAACACTAACAGTCA 839

Qy     781 AGCAAGTCAAAGTAATGACATTGATAGTCTCTTGAAACAGCTCTACAAACTGCCTTTGAG 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     840 AGCAAGTCAAAGTAATGACATTGATAGTCTCTTGAAACAGCTCTACAAACTGCCTTTGAG 899

Qy     841 TCAACGACATGTAGAATCTGATGGCCTTGTCTTTGATCCAGCACAAATCACAAGTCGAAC 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     900 TCAACGACATGTAGAATCTGATGGCCTTGTCTTTGATCCAGCACAAATCACAAGTCGAAC 959

Qy     901 AGCTAGAGGTGTTGCAGTGCCACACGGAGATCATTACCACTTCATCCCTTACTCTCAAAT 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     960 AGCTAGAGGTGTTGCAGTGCCACACGGAGATCATTACCACTTCATCCCTTACTCTCAAAT 1019

Qy     961 GTCTGAATTGGAAGAACGAATCGCTCGTATTATTCCCCTTCGTTATCGTTCAAACCATTG 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1020 GTCTGAATTGGAAGAACGAATCGCTCGTATTATTCCCCTTCGTTATCGTTCAAACCATTG 1079

Qy    1021 GGTACCAGATTCAAGGCCAGAACAACCAAGTCCACAACCGACTCCGGAACCTAGTCCAGG 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1080 GGTACCAGATTCAAGGCCAGAACAACCAAGTCCACAACCGACTCCGGAACCTAGTCCAGG 1139

Qy    1081 CCCGCAACCTGCACCAAATCTTAAAATAGACTCAAATTCTTCTTTGGTTAGTCAGCTGGT 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1140 CCCGCAACCTGCACCAAATCTTAAAATAGACTCAAATTCTTCTTTGGTTAGTCAGCTGGT 1199

Qy    1141 ACGAAAAGTTGGGGAAGGATATGTATTCGAAGAAAAGGGCATCTCTCGTTATGTCTTTGC 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1200 ACGAAAAGTTGGGGAAGGATATGTATTCGAAGAAAAGGGCATCTCTCGTTATGTCTTTGC 1259

Qy    1201 GAAAGATTTACCATCTGAAACTGTTAAAAATCTTGAAAGCAAGTTATCAAAACAAGAGAG 1260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1260 GAAAGATTTACCATCTGAAACTGTTAAAAATCTTGAAAGCAAGTTATCAAAACAAGAGAG 1319

Qy    1261 TGTTTCACACACTTTAACTGCTAAAAAAGAAAATGTTGCTCCTCGTGACCAAGAATTTTA 1320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1320 TGTTTCACACACTTTAACTGCTAAAAAAGAAAATGTTGCTCCTCGTGACCAAGAATTTTA 1379

Qy    1321 TGATAAAGCATATAATCTGTTAACTGAGGCTCATAAAGCCTTGTTTGNAAATAAGGGTCG 1380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1380 TGATAAAGCATATAATCTGTTAACTGAGGCTCATAAAGCCTTGTTTGNAAATAAGGGTCG 1439

Qy    1381 TAATTCTGATTTCCAAGCCTTAGACAAATTATTAGAACGCTTGAATGATGAATCGACTAA 1440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1440 TAATTCTGATTTCCAAGCCTTAGACAAATTATTAGAACGCTTGAATGATGAATCGACTAA 1499

Qy    1441 TAAAGAAAAATTGGTAGATGATTTATTGGCATTCCTAGCACCAATTACCCATCCAGAGCG 1500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1500 TAAAGAAAAATTGGTAGATGATTTATTGGCATTCCTAGCACCAATTACCCATCCAGAGCG 1559

Qy    1501 ACTTGGCAAACCAAATTCTCAAATTGAGTATACTGAAGACGAAGTTCGTATTGCTCAATT 1560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1560 ACTTGGCAAACCAAATTCTCAAATTGAGTATACTGAAGACGAAGTTCGTATTGCTCAATT 1619

Qy    1561 AGCTGATAAGTATACAACGTCAGATGGTTACATTTTTGATGAACATGATATAATCAGTGA 1620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1620 AGCTGATAAGTATACAACGTCAGATGGTTACATTTTTGATGAACATGATATAATCAGTGA 1679
```

```
Qy   1621 TGAAGGAGATGCATATGTAACGCCTCATATGGGCCATAGTCACTGGATTGGAAAAGATAG 1680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1680 TGAAGGAGATGCATATGTAACGCCTCATATGGGCCATAGTCACTGGATTGGAAAAGATAG 1739

Qy   1681 CCTTTCTGATAAGGAAAAAGTTGCAGCTCAAGCCTATACTAAAGAAAAAGGTATCCTACC 1740
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1740 CCTTTCTGATAAGGAAAAAGTTGCAGCTCAAGCCTATACTAAAGAAAAAGGTATCCTACC 1799

Qy   1741 TCCATCTCCAGACGCAGATGTTAAAGCAAATCCAACTGGAGATAGTGCAGCAGCTATTTA 1800
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1800 TCCATCTCCAGACGCAGATGTTAAAGCAAATCCAACTGGAGATAGTGCAGCAGCTATTTA 1859

Qy   1801 CAATCGTGTGAAAGGGGAAAAACGAATTCCACTCGTTCGACTTCCATATATGGTTGAGCA 1860
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1860 CAATCGTGTGAAAGGGGAAAAACGAATTCCACTCGTTCGACTTCCATATATGGTTGAGCA 1919

Qy   1861 TACAGTTGAGGTTAAAAACGGTAATTTGATTATTCCTCATAAGGATCATTACCATAATAT 1920
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1920 TACAGTTGAGGTTAAAAACGGTAATTTGATTATTCCTCATAAGGATCATTACCATAATAT 1979

Qy   1921 TAAATTTGCTTGGTTTGATGATCACACATACAAAGCTCCAAATGGCTATACCTTGGAAGA 1980
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1980 TAAATTTGCTTGGTTTGATGATCACACATACAAAGCTCCAAATGGCTATACCTTGGAAGA 2039

Qy   1981 TTTGTTTGCGACGATTAAGTACTACGTAGAACACCCTGACGAACGTCCACATTCTAATGA 2040
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2040 TTTGTTTGCGACGATTAAGTACTACGTAGAACACCCTGACGAACGTCCACATTCTAATGA 2099

Qy   2041 TGGATGGGGCAATGCCAGTGAGCATGTGTTAGGCAAGAAAGACCACAGTGAAGATCCAAA 2100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2100 TGGATGGGGCAATGCCAGTGAGCATGTGTTAGGCAAGAAAGACCACAGTGAAGATCCAAA 2159

Qy   2101 TAAGAACTTCAAAGCGGATGAAGAGCCAGTAGAGGAAACACCTGCTGAGCCAGAAGTCCC 2160
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2160 TAAGAACTTCAAAGCGGATGAAGAGCCAGTAGAGGAAACACCTGCTGAGCCAGAAGTCCC 2219

Qy   2161 TCAAGTAGAGACTGAAAAAGTAGAAGCCCAACTCAAAGAAGCAGAAGTTTTGCTTGCGAA 2220
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2220 TCAAGTAGAGACTGAAAAAGTAGAAGCCCAACTCAAAGAAGCAGAAGTTTTGCTTGCGAA 2279

Qy   2221 AGTAACGGATTCTAGTCTGAAAGCCAATGCAACAGAAACTCTAGCTGGTTTACGAAATAA 2280
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2280 AGTAACGGATTCTAGTCTGAAAGCCAATGCAACAGAAACTCTAGCTGGTTTACGAAATAA 2339

Qy   2281 TTTGACTCTTCAAATTATGGATAACAATAGTATCATGGCAGAAGCAGAAAAATTACTTGC 2340
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2340 TTTGACTCTTCAAATTATGGATAACAATAGTATCATGGCAGAAGCAGAAAAATTACTTGC 2399

Qy   2341 GTTGTTAAAAGGAAGTAATCCTTCATCTGTAAGTAAGGAAAAAATAAAC 2389
          ||||||||||||||||||||||||||||||||||||||||||||||||
Db   2400 GTTGTTAAAAGGAAGTAATCCTTCATCTGTAAGTAAGGAAAAAATAAAC 2448
```

RESULT 4
AAV52227
ID AAV52227 standard; DNA; 8195 BP.
XX
AC AAV52227;
XX
DT 23-OCT-1998 (first entry)
XX
DE Streptococcus pneumoniae genome fragment SEQ ID NO:94.
XX
KW Streptococcus pneumoniae; S. pneumoniae; genome; diagnosis; assay;
KW computer readable medium; vaccine; pharmaceutical composition; ds.
XX
OS Streptococcus pneumoniae.
XX
PN WO9818931-A2.
XX
PD 07-MAY-1998.
XX
PF 30-OCT-1997; 97WO-US19588.
XX
PR 31-OCT-1996; 96US-0029960.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Barash SC, Choi GH, Dillon PJ, Dougherty BA, Fannon M;
PI Kunsch CA, Rosen CA;
XX
DR WPI; 1998-272225/24.
XX
PT Computer-readable medium with recorded Streptococcus pneumoniae
PT polynucleotide sequences - useful in diagnostic kits and assays, and
PT pharmaceutical compositions and vaccines for Streptococcus
PT pneumoniae
XX
PS Claim 1; Page 727-732; 1409pp; English.
XX
CC The present invention describes a computer readable medium which has
CC the nucleotide sequences SEQ ID NO:1 to 391 (AAV52134 to AAV52524)
CC recorded on it, or a representative fragment or a sequence at least 95%
CC identical to SEQ ID NO: 1 to 391. The nucleotide sequences depicted in
CC SEQ ID NO:1 to 391 (AAV52134 to AAV52524) are genomic fragments from
CC Streptococcus pneumoniae. The present invention also describes an
CC isolated nucleic acid molecule encoding a homologue of any of the
CC fragments of the S.pneumoniae genome (SEQ ID NO:1 to 391) where the
CC nucleic acid molecule is produced by a process comprising: (a) screening
CC a genomic DNA library using as a probe a target sequence defined by any
CC of the sequences in SEQ ID NO:1 to 391, identifying members of the
CC library which contain sequences that hybridise to the target sequence and
CC isolating the nucleic acid molecules from the members; or (b) isolating
CC mRNA, DNA or cDNA produced from an organism, amplifying nucleic acid
CC molecules whose nucleotide sequence is homologous to amplification
CC primers derived from the fragment of the S. pneumoniae genome to prime
CC the amplification and isolating the amplified sequences. The computer
CC readable medium can be used in a computer-based system for identifying
CC fragments of the S. pneumoniae genome of commercial importance, or
CC expression modulating fragments of the S. pneumoniae genome. Products
CC from the present invention can be used in diagnosis kits and assays, and
CC pharmaceutical compositions and vaccines for S. pneumoniae.
XX
SQ Sequence 8195 BP; 2688 A; 1622 C; 1777 G; 2105 T; 3 other;

Query Match 100.0%; Score 2388; DB 19; Length 8195;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2388; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy      1 TTCTTACGAGTTGGGACTGTATCAAGCTAGAACGGTTAAGGAAAATAATCGTGTTTCCTA 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3053 TTCTTACGAGTTGGGACTGTATCAAGCTAGAACGGTTAAGGAAAATAATCGTGTTTCCTA 3112

Qy     61 TATAGATGGAAAACAAGCGACGCAAAAAACGGAGAATTTGACTCCTGATGAGGTTAGCAA 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3113 TATAGATGGAAAACAAGCGACGCAAAAAACGGAGAATTTGACTCCTGATGAGGTTAGCAA 3172

Qy    121 GCGTGAAGGAATCAATGCTGAGCAAATCGTCATCAAGATAACAGACCAAGGCTATGTCAC 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3173 GCGTGAAGGAATCAATGCTGAGCAAATCGTCATCAAGATAACAGACCAAGGCTATGTCAC 3232

Qy    181 TTCACATGGCGACCACTATCATTATTACAATGGTAAGGTTCCTTATGACGCTATCATCAG 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3233 TTCACATGGCGACCACTATCATTATTACAATGGTAAGGTTCCTTATGACGCTATCATCAG 3292

Qy    241 TGAAGAATTACTCATGAAAGATCCAAACTATAAGCTAAAAGATGAGGATATTGTTAATGA 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3293 TGAAGAATTACTCATGAAAGATCCAAACTATAAGCTAAAAGATGAGGATATTGTTAATGA 3352

Qy    301 GGTCAAGGGTGGATATGTTATCAAGGTAGATGGAAAATACTATGTTTACCTTAAGGATGC 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3353 GGTCAAGGGTGGATATGTTATCAAGGTAGATGGAAAATACTATGTTTACCTTAAGGATGC 3412

Qy    361 TGCCCACGCGGATAACGTCCGTACAAAAGAGGAAATCAATCGACAAAAACAAGAGCATAG 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db   3413 TGCCCACGCGGATAACGTCCGTACAAAAGAGGAAATCAATCGACAAAAACAAGAGCATAG 3472

Qy    421 TCAACATCGTGAAGGTGGAACTCCAAGAAACGATGGTGCTGTTGCCTTGGCACGTTCGCA 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3473 TCAACATCGTGAAGGTGGAACTCCAAGAAACGATGGTGCTGTTGCCTTGGCACGTTCGCA 3532

Qy    481 AGGACGCTATACTACAGATGATGGTTATATCTTTAATGCTTCTGATATCATAGAGGATAC 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3533 AGGACGCTATACTACAGATGATGGTTATATCTTTAATGCTTCTGATATCATAGAGGATAC 3592

Qy    541 TGGTGATGCTTATATCGTTCCTCATGGAGATCATTACCATTACATTCCTAAGAATGAGTT 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3593 TGGTGATGCTTATATCGTTCCTCATGGAGATCATTACCATTACATTCCTAAGAATGAGTT 3652

Qy    601 ATCAGCTAGCGAGTTGGCTGCTGCAGAAGCCTTCCTATCTGGTCGAGGAAATCTGTCAAA 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3653 ATCAGCTAGCGAGTTGGCTGCTGCAGAAGCCTTCCTATCTGGTCGAGGAAATCTGTCAAA 3712

Qy    661 TTCAAGAACCTATCGCCGACAAAATAGCGATAACACTTCAAGAACAAACTGGGTACCTTC 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3713 TTCAAGAACCTATCGCCGACAAAATAGCGATAACACTTCAAGAACAAACTGGGTACCTTC 3772

Qy    721 TGTAAGCAATCCAGGAACTACAAATACTAACACAAGCAACAACAGCAACACTAACAGTCA 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3773 TGTAAGCAATCCAGGAACTACAAATACTAACACAAGCAACAACAGCAACACTAACAGTCA 3832

Qy    781 AGCAAGTCAAAGTAATGACATTGATAGTCTCTTGAAACAGCTCTACAAACTGCCTTTGAG 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3833 AGCAAGTCAAAGTAATGACATTGATAGTCTCTTGAAACAGCTCTACAAACTGCCTTTGAG 3892

Qy    841 TCAACGACATGTAGAATCTGATGGCCTTGTCTTTGATCCAGCACAAATCACAAGTCGAAC 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3893 TCAACGACATGTAGAATCTGATGGCCTTGTCTTTGATCCAGCACAAATCACAAGTCGAAC 3952

Qy    901 AGCTAGAGGTGTTGCAGTGCCACACGGAGATCATTACCACTTCATCCCTTACTCTCAAAT 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3953 AGCTAGAGGTGTTGCAGTGCCACACGGAGATCATTACCACTTCATCCCTTACTCTCAAAT 4012

Qy    961 GTCTGAATTGGAAGAACGAATCGCTCGTATTATTCCCCTTCGTTATCGTTCAAACCATTG 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4013 GTCTGAATTGGAAGAACGAATCGCTCGTATTATTCCCCTTCGTTATCGTTCAAACCATTG 4072

Qy   1021 GGTACCAGATTCAAGGCCAGAACAACCAAGTCCACAACCGACTCCGGAACCTAGTCCAGG 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4073 GGTACCAGATTCAAGGCCAGAACAACCAAGTCCACAACCGACTCCGGAACCTAGTCCAGG 4132

Qy   1081 CCCGCAACCTGCACCAAATCTTAAAATAGACTCAAATTCTTCTTTGGTTAGTCAGCTGGT 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4133 CCCGCAACCTGCACCAAATCTTAAAATAGACTCAAATTCTTCTTTGGTTAGTCAGCTGGT 4192

Qy   1141 ACGAAAAGTTGGGGAAGGATATGTATTCGAAGAAAAGGGCATCTCTCGTTATGTCTTTGC 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4193 ACGAAAAGTTGGGGAAGGATATGTATTCGAAGAAAAGGGCATCTCTCGTTATGTCTTTGC 4252

Qy   1201 GAAAGATTTACCATCTGAAACTGTTAAAAATCTTGAAAGCAAGTTATCAAAACAAGAGAG 1260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4253 GAAAGATTTACCATCTGAAACTGTTAAAAATCTTGAAAGCAAGTTATCAAAACAAGAGAG 4312

Qy   1261 TGTTTCACACACTTTAACTGCTAAAAAAGAAAATGTTGCTCCTCGTGACCAAGAATTTTA 1320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4313 TGTTTCACACACTTTAACTGCTAAAAAAGAAAATGTTGCTCCTCGTGACCAAGAATTTTA 4372

Qy   1321 TGATAAAGCATATAATCTGTTAACTGAGGCTCATAAAGCCTTGTTTGNAAATAAGGGTCG 1380
          ||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db   4373 TGATAAAGCATATAATCTGTTAACTGAGGCTCATAAAGCCTTGTTTGAAAATAAGGGTCG 4432

Qy   1381 TAATTCTGATTTCCAAGCCTTAGACAAATTATTAGAACGCTTGAATGATGAATCGACTAA 1440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4433 TAATTCTGATTTCCAAGCCTTAGACAAATTATTAGAACGCTTGAATGATGAATCGACTAA 4492

Qy   1441 TAAAGAAAAATTGGTAGATGATTTATTGGCATTCCTAGCACCAATTACCCATCCAGAGCG 1500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4493 TAAAGAAAAATTGGTAGATGATTTATTGGCATTCCTAGCACCAATTACCCATCCAGAGCG 4552

Qy   1501 ACTTGGCAAACCAAATTCTCAAATTGAGTATACTGAAGACGAAGTTCGTATTGCTCAATT 1560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4553 ACTTGGCAAACCAAATTCTCAAATTGAGTATACTGAAGACGAAGTTCGTATTGCTCAATT 4612

Qy   1561 AGCTGATAAGTATACAACGTCAGATGGTTACATTTTTGATGAACATGATATAATCAGTGA 1620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4613 AGCTGATAAGTATACAACGTCAGATGGTTACATTTTTGATGAACATGATATAATCAGTGA 4672

Qy   1621 TGAAGGAGATGCATATGTAACGCCTCATATGGGCCATAGTCACTGGATTGGAAAAGATAG 1680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4673 TGAAGGAGATGCATATGTAACGCCTCATATGGGCCATAGTCACTGGATTGGAAAAGATAG 4732

Qy   1681 CCTTTCTGATAAGGAAAAAGTTGCAGCTCAAGCCTATACTAAAGAAAAAGGTATCCTACC 1740
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4733 CCTTTCTGATAAGGAAAAAGTTGCAGCTCAAGCCTATACTAAAGAAAAAGGTATCCTACC 4792

Qy   1741 TCCATCTCCAGACGCAGATGTTAAAGCAAATCCAACTGGAGATAGTGCAGCAGCTATTTA 1800
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4793 TCCATCTCCAGACGCAGATGTTAAAGCAAATCCAACTGGAGATAGTGCAGCAGCTATTTA 4852

Qy   1801 CAATCGTGTGAAAGGGGAAAAACGAATTCCACTCGTTCGACTTCCATATATGGTTGAGCA 1860
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4853 CAATCGTGTGAAAGGGGAAAAACGAATTCCACTCGTTCGACTTCCATATATGGTTGAGCA 4912

Qy   1861 TACAGTTGAGGTTAAAAACGGTAATTTGATTATTCCTCATAAGGATCATTACCATAATAT 1920
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4913 TACAGTTGAGGTTAAAAACGGTAATTTGATTATTCCTCATAAGGATCATTACCATAATAT 4972

Qy   1921 TAAATTTGCTTGGTTTGATGATCACACATACAAAGCTCCAAATGGCTATACCTTGGAAGA 1980
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4973 TAAATTTGCTTGGTTTGATGATCACACATACAAAGCTCCAAATGGCTATACCTTGGAAGA 5032

Qy   1981 TTTGTTTGCGACGATTAAGTACTACGTAGAACACCCTGACGAACGTCCACATTCTAATGA 2040
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5033 TTTGTTTGCGACGATTAAGTACTACGTAGAACACCCTGACGAACGTCCACATTCTAATGA 5092

Qy   2041 TGGATGGGGCAATGCCAGTGAGCATGTGTTAGGCAAGAAAGACCACAGTGAAGATCCAAA 2100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5093 TGGATGGGGCAATGCCAGTGAGCATGTGTTAGGCAAGAAAGACCACAGTGAAGATCCAAA 5152

Qy   2101 TAAGAACTTCAAAGCGGATGAAGAGCCAGTAGAGGAAACACCTGCTGAGCCAGAAGTCCC 2160
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5153 TAAGAACTTCAAAGCGGATGAAGAGCCAGTAGAGGAAACACCTGCTGAGCCAGAAGTCCC 5212

Qy   2161 TCAAGTAGAGACTGAAAAAGTAGAAGCCCAACTCAAAGAAGCAGAAGTTTTGCTTGCGAA 2220
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5213 TCAAGTAGAGACTGAAAAAGTAGAAGCCCAACTCAAAGAAGCAGAAGTTTTGCTTGCGAA 5272

Qy   2221 AGTAACGGATTCTAGTCTGAAAGCCAATGCAACAGAAACTCTAGCTGGTTTACGAAATAA 2280
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5273 AGTAACGGATTCTAGTCTGAAAGCCAATGCAACAGAAACTCTAGCTGGTTTACGAAATAA 5332

Qy   2281 TTTGACTCTTCAAATTATGGATAACAATAGTATCATGGCAGAAGCAGAAAAATTACTTGC 2340
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5333 TTTGACTCTTCAAATTATGGATAACAATAGTATCATGGCAGAAGCAGAAAAATTACTTGC 5392

Qy   2341 GTTGTTAAAAGGAAGTAATCCTTCATCTGTAAGTAAGGAAAAAATAAAC 2389
          |||||||||||||||||||||||||||||||||||||||||||||||||
Db   5393 GTTGTTAAAAGGAAGTAATCCTTCATCTGTAAGTAAGGAAAAAATAAAC 5441
```

```
RESULT 5
AAA65731
ID   AAA65731 standard; DNA; 2523 BP.
XX
AC   AAA65731;
XX
DT   21-NOV-2000  (first entry)
XX
DE   Streptococcus pneumoniae BVH-11 gene SEQ ID NO:3.
XX
KW   Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;
```

KW prophylaxis; therapy; infection; diagnosis; meningitis; bacteraemia;
KW otitis media; pneumonia; immunisation; bactericidal; ds.
XX
OS Streptococcus pneumoniae.
XX
PN WO200039299-A2.
XX
PD 06-JUL-2000.
XX
PF 20-DEC-1999; 99WO-CA01218.
XX
PR 23-DEC-1998; 98US-0113800.
XX
PA (BIOC-) BIOCHEM PHARMA INC.
XX
PI Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N;
XX
DR WPI; 2000-452397/39.
DR P-PSDB; AAB12716.
XX
PT Streptococcal antigens useful for vaccinating against e.g. meningitis,
PT otitis media, bacteremia and/or pneumonia -
XX
PS Example 2; Fig 3; 106pp; English.
XX
CC The present invention describes nucleic acids (I) encoding protein
CC antigens (II) from Streptococcus pneumoniae. The protein antigens
CC have bactericidal activity. The nucleic acids, encoding the protein
CC antigens, may be used for the recombinant production of the proteins
CC they encode. The protein antigens may then be used as vaccines for the
CC prevention and treatment of Streptococcal infections in mammals
CC (especially humans) which result in, e.g. meningitis, otitis media,
CC bacteraemia and/or pneumonia. The present sequence encodes the
CC S. pneumoniae BVH-11 protein antigen.
XX
SQ Sequence 2523 BP; 879 A; 523 C; 526 G; 595 T; 0 other;

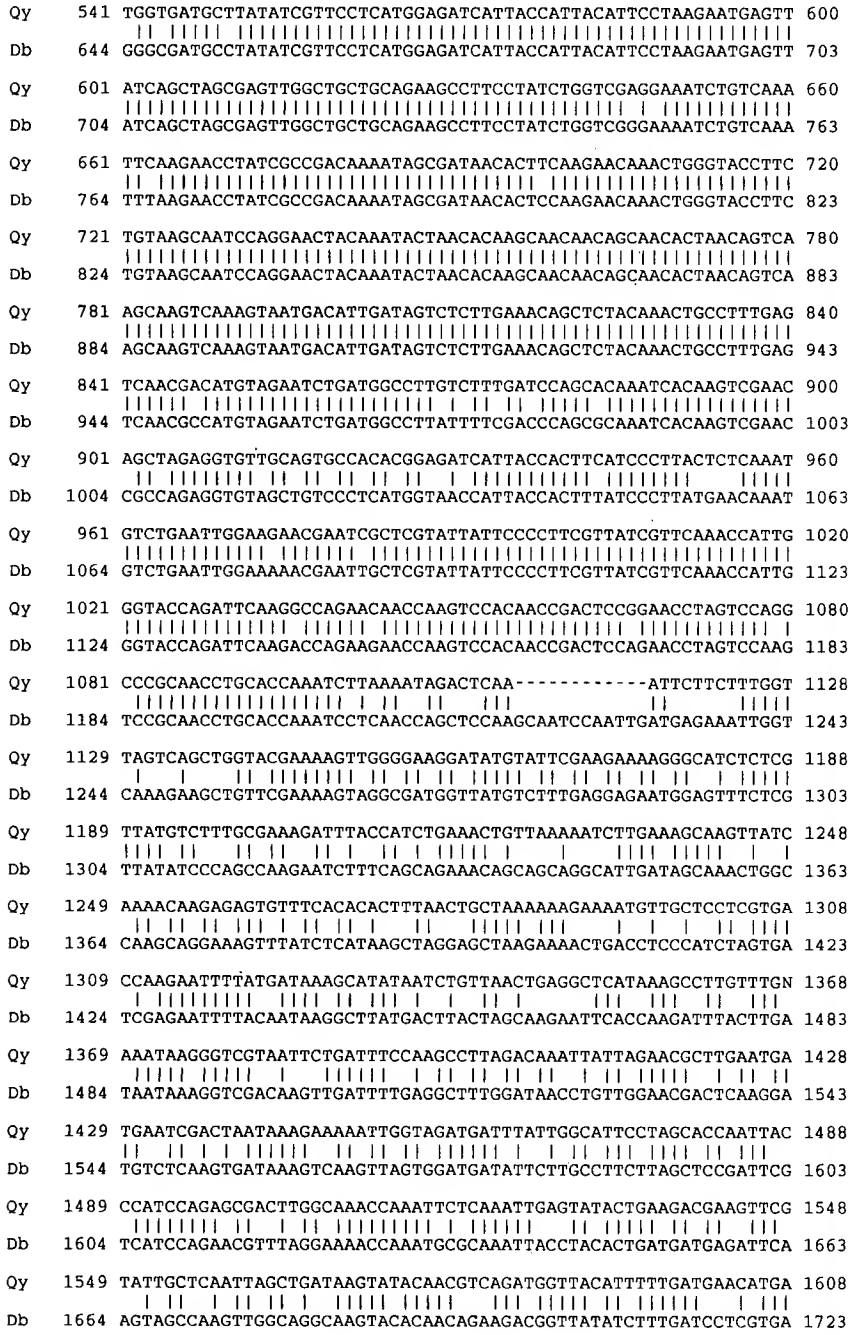
Query Match 57.5%; Score 1374.2; DB 21; Length 2523;
Best Local Similarity 73.9%; Pred. No. 0;
Matches 1819; Conservative 0; Mismatches 539; Indels 102; Gaps 2;

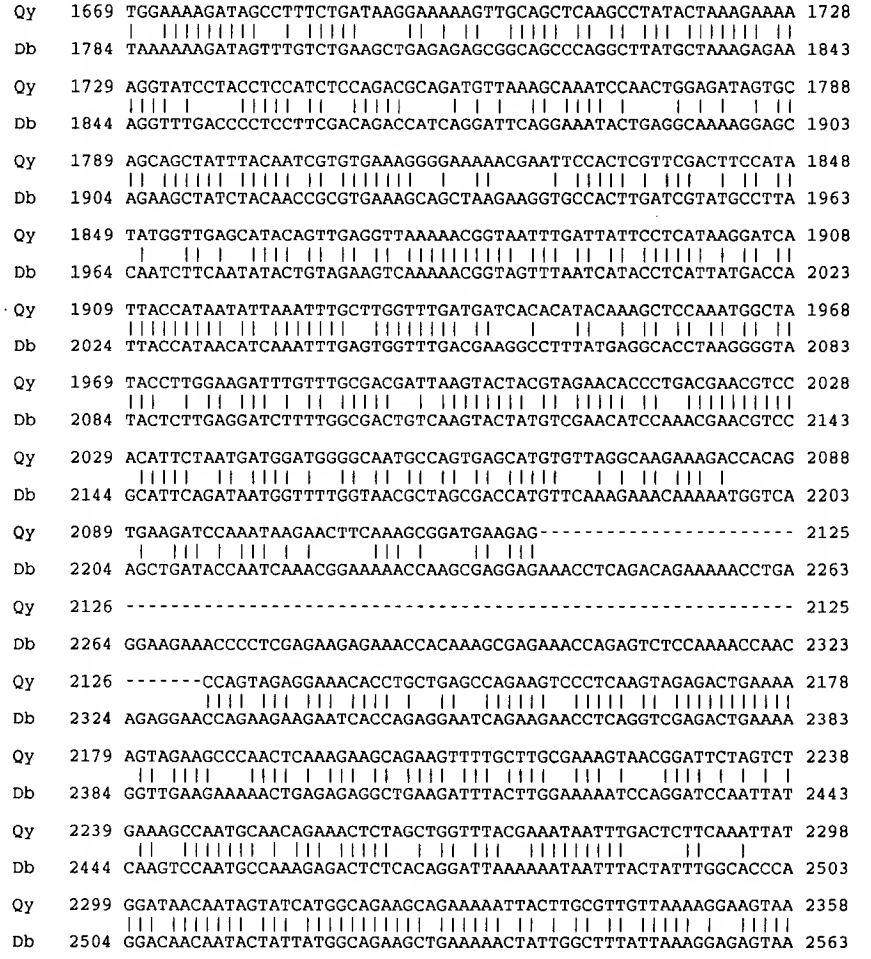
```
Db   1620 AGTAGCCAAGTTGGCAGGCAAGTACACAACAGAAGACGGTTATATCTTTGATCCTCGTGA 1679

Qy   1609 TATAATCAGTGATGAAGGAGATGCATATGTAACGCCTCATATGGGCCATAGTCACTGGAT 1668
Db   1680 TATAACCAGTGATGAGGGGGATGCCTATGTAACTCCACATATGACCCATAGCCACTGGAT 1739

Qy   1669 TGGAAAAGATAGCCTTTCTGATAAGGAAAAAGTTGCAGCTCAAGCCTATACTAAAGAAAA 1728
Db   1740 TAAAAAAGATAGTTTGTCTGAAGCTGAGAGAGCGGCAGCCCAGGCTTATGCTAAAGAGAA 1799

Qy   1729 AGGTATCCTACCTCCATCTCCAGACGCAGATGTTAAAGCAAATCCAACTGGAGATAGTGC 1788
Db   1800 AGGTTTGACCCCTCCTTCGACAGACCATCAGGATTCAGGAAATACTGAGGCAAAAGGAGC 1859

Qy   1789 AGCAGCTATTTACAATCGTGTGAAAGGGGAAAAACGAATTCCACTCGTTCGACTTCCATA 1848
Db   1860 AGAAGCTATCTACAACCGCGTGAAAGCAGCTAAGAAGGTGCCACTTGATCGTATGCCTTA 1919

Qy   1849 TATGGTTGAGCATACAGTTGAGGTTAAAAACGGTAATTTGATTATTCCTCATAAGGATCA 1908
Db   1920 CAATCTTCAATATACTGTAGAAGTCAAAAACGGTAGTTTAATCATACCTCATTATGACCA 1979

Qy   1909 TTACCATAATATTAAATTTGCTTGGTTTGATGATCACACATACAAAGCTCCAAATGGCTA 1968
Db   1980 TTACCATAACATCAAATTTGAGTGGTTTGACGAAGGCCTTTATGAGGCACCTAAGGGGTA 2039

Qy   1969 TACCTTGGAAGATTTGTTTGCGACGATTAAGTACTACGTAGAACACCCTGACGAACGTCC 2028
Db   2040 TACTCTTGAGGATCTTTTGGCGACTGTCAAGTACTATGTCGAACATCCAAACGAACGTCC 2099

Qy   2029 ACATTCTAATGATGGATGGGGCAATGCCAGTGAGCATGTGTTAGGCAAGAAAGACCACAG 2088
Db   2100 GCATTCAGATAATGGTTTTGGTAACGCTAGCGACCATGTTCAAAGAAACAAAAATGGTCA 2159

Qy   2089 TGAAGATCCAAATAAGAACTTCAAAGCGGATGAAGAG---------------------- 2125
Db   2160 AGCTGATACCAATCAAACGGAAAAACCAAGCGAGGAGAAACCTCAGACAGAAAAACCTGA 2219

Qy   2126 ------------------------------------------------------------ 2125
Db   2220 GGAAGAAACCCCTCGAGAAGAGAAACCACAAAGCGAGAAACCAGAGTCTCCAAAACCAAC 2279

Qy   2126 -------CCAGTAGAGGAAACACCTGCTGAGCCAGAAGTCCCTCAAGTAGAGACTGAAAA 2178
Db   2280 AGAGGAACCAGAAGAAGAATCACCAGAGGAATCAGAAGAACCTCAGGTCGAGACTGAAAA 2339

Qy   2179 AGTAGAAGCCCAACTCAAAGAAGCAGAAGTTTTGCTTGCGAAAGTAACGGATTCTAGTCT 2238
Db   2340 GGTTGAAGAAAAACTGAGAGAGGCTGAAGATTTACTTGGAAAAATCCAGGATCCAATTAT 2399

Qy   2239 GAAAGCCAATGCAACAGAAACTCTAGCTGGTTTACGAAATAATTTGACTCTTCAAATTAT 2298
Db   2400 CAAGTCCAATGCCAAAGAGACTCTCACAGGATTAAAAAATAATTTACTATTTGGCACCCA 2459

Qy   2299 GGATAACAATAGTATCATGGCAGAAGCAGAAAAATTACTTGCGTTGTTAAAAGGAAGTAA 2358
Db   2460 GGACAACAATACTATTATGGCAGAAGCTGAAAAACTATTGGCTTTATTAAAGGAGAGTAA 2519
```

RESULT 6
AAA65736
ID AAA65736 standard; DNA; 2647 BP.
XX
AC AAA65736;
XX
DT 21-NOV-2000 (first entry)
XX
DE Streptococcus pneumoniae BVH-11 gene SEQ ID NO:12.
XX
KW Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;
KW prophylaxis; therapy; infection; diagnosis; meningitis; bacteraemia;
KW otitis media; pneumonia; immunisation; bactericidal; ds.
XX
OS Streptococcus pneumoniae.
XX
PN WO200039299-A2.
XX
PD 06-JUL-2000.
XX
PF 20-DEC-1999; 99WO-CA01218.
XX
PR 23-DEC-1998; 98US-0113800.
XX
PA (BIOC-) BIOCHEM PHARMA INC.
XX
PI Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N;
XX
DR WPI; 2000-452397/39.
XX
PT Streptococcal antigens useful for vaccinating against e.g. meningitis,
PT otitis media, bacteremia and/or pneumonia -
XX
PS Example 6; Fig 15; 106pp; English.
XX
CC The present invention describes nucleic acids (I) encoding protein
CC antigens (II) from Streptococcus pneumoniae. The protein antigens
CC have bactericidal activity. The nucleic acids, encoding the protein
CC antigens, may be used for the recombinant production of the proteins
CC they encode. The protein antigens may then be used as vaccines for the
CC prevention and treatment of Streptococcal infections in mammals
CC (especially humans) which result in, e.g. meningitis, otitis media,
CC bacteraemia and/or pneumonia. The present sequence encodes the
CC S. pneumoniae BVH-11 protein antigen.
XX
SQ Sequence 2647 BP; 934 A; 538 C; 556 G; 619 T; 0 other;

Query Match 57.5%; Score 1374.2; DB 21; Length 2647;
Best Local Similarity 73.9%; Pred. No. 0;
Matches 1819; Conservative 0; Mismatches 539; Indels 102; Gaps 2;

```
Qy      1 TTCTTACGAGTTGGGACTGTATCAAGCTAGAACGGTTAAGGAAAATAATCGTGTTTCCTA 60
Db    104 TGCTTATGAACTAGGTTTGCATCAAGCTCAAACTGTAAAAGAAAATAATCGTGTTTCCTA 163

Qy     61 TATAGATGGAAAACAAGCGACGCAAAAAACGGAGAATTTGACTCCTGATGAGGTTAGCAA 120
Db    164 TATAGATGGAAAACAAGCGACGCAAAAAACGGAGAATTTGACTCCTGATGAGGTTAGCAA 223

Qy    121 GCGTGAAGGAATCAATGCTGAGCAAATCGTCATCAAGATAACAGACCAAGGCTATGTCAC 180
Db    224 GCGTGAAGGAATCAACGCCGAACAAATCGTCATCAAGATTACGGATCAAGGTTATGTGAC 283

Qy    181 TTCACATGGCGACCACTATCATTATTACAATGGTAAGGTTCCTTATGACGCTATCATCAG 240
Db    284 CTCTCATGGAGACCATTATCATTACTATAATGGCAAGGTCCCTTATGATGCCATCATCAG 343

Qy    241 TGAAGAATTACTCATGAAAGATCCAAACTATAAGCTAAAAGATGAGGATATTGTTAATGA 300
Db    344 TGAAGAGCTCCTCATGAAAGATCCGAATTATCAGTTGAAGGATTCAGACATTGTCAATGA 403

Qy    301 GGTCAAGGGTGGATATGTTATCAAGGTAGATGGAAAATACTATGTTTACCTTAAGGATGC 360
Db    404 AATCAAGGGTGGTTATGTCATTAAGGTAAACGGTAAATACTATGTTTACCTTAAGGATGC 463

Qy    361 TGCCCACGCGGATAACGTCCGTACAAAAGAGGAAATCAATCGACAAAAACAAGAGCATAG 420
Db    464 AGCTCATGCGGATAATGTCCGTACAAAAGAAGAAATCAATCGGCAAAAACAAGAACATAG 523

Qy    421 TCAACATCGTGAAGGTGGAACTCCAAGAAACGATGGTGCTGTTGCCTTGGCACGTTCGCA 480
Db    524 TCAGCATCGTGAAGGAGGGACTTCAGCAAACGATGGTGCGGTAGCCTTTGCACGTTCACA 583

Qy    481 AGGACGCTATACTACAGATGATGGTTATATCTTTAATGCTTCTGATATCATAGAGGATAC 540
Db    584 GGGACGCTACACCACAGATGATGGTTATATCTTCAATGCATCTGATATCATCGAAGATAC 643
```





RESULT 7
ABK15103
ID ABK15103 standard; DNA; 2647 BP.
XX
AC ABK15103;
XX
DT 08-MAY-2002 (first entry)
XX
DE DNA encoding Streptococcus pneumoniae BVH-11.
XX
KW BVH-3; BVH-11; vaccine; meningitis; otitis media; bacteraemia;
KW pneumonia; streptococcal bacterial infection; gene; ds.
XX
OS Streptococcus pneumoniae.
XX
FH Key Location/Qualifiers

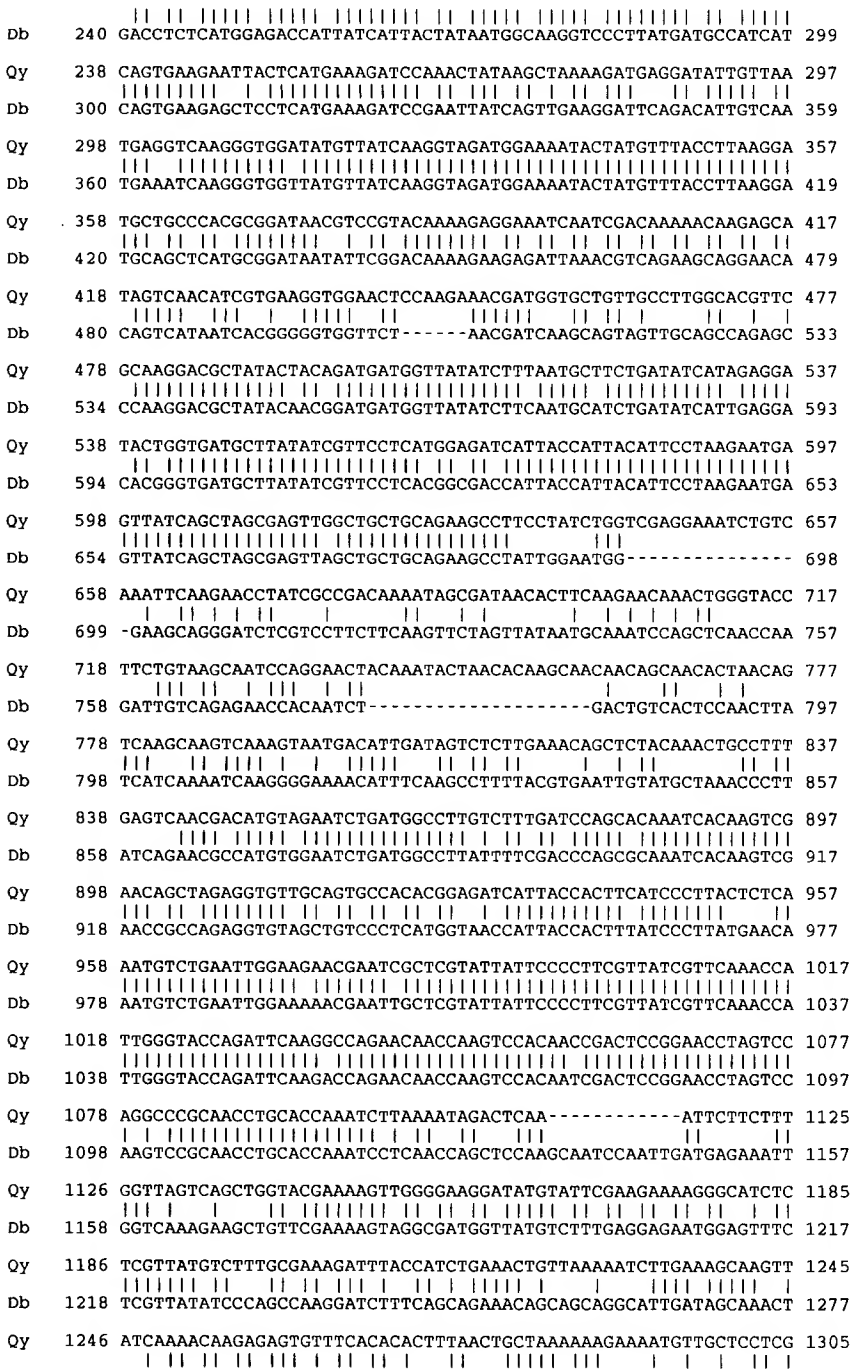
```
FT   CDS             45..2567
FT                   /*tag=  a
FT                   /product= "BVH-11"
FT                   /note= "The gene is flanked by sequences from the
FT                          vector SP64, no information on which is
FT                          given in the specification"
XX
PN   WO200198334-A2.
XX
PD   27-DEC-2001.
XX
PF   19-JUN-2001; 2001WO-CA00908.
XX
PR   20-JUN-2000; 2000US-212683P.
XX
PA   (SHIR-) SHIRE BIOCHEM INC.
XX
PI   Hamel J,  Ouellet C,  Charland N,  Martin D,  Brodeur B;
XX
DR   WPI; 2002-122272/16.
DR   P-PSDB; AAU75933.
XX
PT   New Streptococcus pneumoniae BVH-3 and BVH-11 variant and
PT   epitope-bearing polypeptides, useful as vaccine components for treating
PT   or preventing streptococcal infections such as otitis media,
PT   meningitis, and bacteraemia  -
XX
PS   Disclosure; Fig 4; 113pp; English.
XX
CC   The invention describes an isolated polypeptide (I) with 70-90%
CC   identity to Streptococcus pneumonia protein BVH-3, BVH-11, variants of
CC   BVH-3 or BVH-11, or chimeric sequences derived from them. A vaccine (II)
CC   comprising (I) is useful for therapeutic or prophylactic treatment of
CC   meningitis, otitis media, bacteraemia or pneumonia infection in an
CC   individual susceptible to these disorders. (II) is also useful for
CC   therapeutic or prophylactic treatment of any streptococcal bacterial
CC   infection (e.g., caused by Streptococcus pneumoniae, group A
CC   Streptococcus such as Streptococcus pyogenes, group B Streptococcus such
CC   as Streptococcus agalactiae, S. dysgalactiae, S. uberis, S. nocardia or
CC   Staphylococcus aureus) in an individual susceptible to the infection.
CC   A polynucleotide (III) encoding (I) is useful in DNA immunisation
CC   techniques. The Streptococcus polypeptides are useful in a diagnostic
CC   test for S. pneumoniae infection. (III) is useful for designing DNA
CC   probes for use in detecting the presence of Streptococcus in a biological
CC   sample suspected of containing the bacteria. The DNA probes may also be
CC   used for detecting circulating S. pneumonia nucleic acid in a sample for
CC   diagnosing streptococcal infections. This sequence encodes the
CC   Streptococcus pneumoniae protein BVH-11, used to create the antigenic
CC   peptides described in the method of the invention.
XX
SQ   Sequence 2647 BP; 934 A; 538 C; 556 G; 619 T; 0 other;

  Query Match             57.5%;  Score 1374.2;  DB 24;  Length 2647;
  Best Local Similarity   73.9%;  Pred. No. 0;
  Matches 1819;  Conservative    0;  Mismatches  539;  Indels  102;  Gaps    2;

Qy      1 TTCTTACGAGTTGGGACTGTATCAAGCTAGAACGGTTAAGGAAAATAATCGTGTTTCCTA 60
          | ||||  ||  | ||  || ||||||||  ||| || || ||||||||||||||||||||
Db    104 TGCTTATGAACTAGGTTTGCATCAAGCTCAAACTGTAAAAGAAAATAATCGTGTTTCCTA 163

Qy     61 TATAGATGGAAAACAAGCGACGCAAAAAACGGAGAATTTGACTCCTGATGAGGTTAGCAA 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    164 TATAGATGGAAAACAAGCGACGCAAAAAACGGAGAATTTGACTCCTGATGAGGTTAGCAA 223

Qy    121 GCGTGAAGGAATCAATGCTGAGCAAATCGTCATCAAGATAACAGACCAAGGCTATGTCAC 180
          |||||||||||||||| || || |||||||||||||||| || || ||||| ||||| ||
Db    224 GCGTGAAGGAATCAACGCCGAACAAATCGTCATCAAGATTACGGATCAAGGTTATGTGAC 283

Qy    181 TTCACATGGCGACCACTATCATTATTACAATGGTAAGGTTCCTTATGACGCTATCATCAG 240
          || ||||| ||||| |||||||| || ||||| ||||| |||||||| || ||||||||
Db    284 CTCTCATGGAGACCATTATCATTACTATAATGGCAAGGTCCCTTATGATGCCATCATCAG 343

Qy    241 TGAAGAATTACTCATGAAAGATCCAAACTATAAGCTAAAAGATGAGGATATTGTTAATGA 300
          ||||||  | ||||||||||||||| || ||| || | || |||   || ||||| ||||||
Db    344 TGAAGAGCTCCTCATGAAAGATCCGAATTATCAGTTGAAGGATTCAGACATTGTCAATGA 403

Qy    301 GGTCAAGGGTGGATATGTTATCAAGGTAGATGGAAAATACTATGTTTACCTTAAGGATGC 360
          ||||||||| ||||| || |||||| | || ||||||||||||||||||||||||||||||
Db    404 AATCAAGGGTGGTTATGTCATTAAGGTAAACGGTAAATACTATGTTTACCTTAAGGATGC 463

Qy    361 TGCCCACGCGGATAACGTCCGTACAAAAGAGGAAATCAATCGACAAAAACAAGAGCATAG 420
          || || |||||||| |||||||||||||| |||||||||| |||||||||| ||||||
Db    464 AGCTCATGCGGATAATGTCCGTACAAAAGAAGAAATCAATCGGCAAAAACAAGAACATAG 523

Qy    421 TCAACATCGTGAAGGTGGAACTCCAAGAAACGATGGTGCTGTTGCCTTGGCACGTTCGCA 480
          ||| |||||||||||| || ||| ||  |||||||||||| || ||||| ||||||||||
Db    524 TCAGCATCGTGAAGGAGGGACTTCAGCAAACGATGGTGCGGTAGCCTTTGCACGTTCACA 583

Qy    481 AGGACGCTATACTACAGATGATGGTTATATCTTTAATGCTTCTGATATCATAGAGGATAC 540
          |||||||| || |||||||||||||||||||| ||||| |||||||||| || ||||||
Db    584 GGGACGCTACACCACAGATGATGGTTATATCTTCAATGCATCTGATATCATCGAAGATAC 643

Qy    541 TGGTGATGCTTATATCGTTCCTCATGGAGATCATTACCATTACATTCCTAAGAATGAGTT 600
          || ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    644 GGGCGATGCCTATATCGTTCCTCATGGAGATCATTACCATTACATTCCTAAGAATGAGTT 703

Qy    601 ATCAGCTAGCGAGTTGGCTGCTGCAGAAGCCTTCCTATCTGGTCGAGGAAATCTGTCAAA 660
          |||||||||||||||||||||||||||||||||||||||||||| | ||||||||||||
Db    704 ATCAGCTAGCGAGTTGGCTGCTGCAGAAGCCTTCCTATCTGGTCGGGAAAATCTGTCAAA 763

Qy    661 TTCAAGAACCTATCGCCGACAAAATAGCGATAACACTTCAAGAACAAACTGGGTACCTTC 720
          || |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db    764 TTTAAGAACCTATCGCCGACAAAATAGCGATAACACTCCAAGAACAAACTGGGTACCTTC 823

Qy    721 TGTAAGCAATCCAGGAACTACAAATACTAACACAAGCAACAACAGCAACACTAACAGTCA 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    824 TGTAAGCAATCCAGGAACTACAAATACTAACACAAGCAACAACAGCAACACTAACAGTCA 883

Qy    781 AGCAAGTCAAAGTAATGACATTGATAGTCTCTTGAAACAGCTCTACAAACTGCCTTTGAG 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    884 AGCAAGTCAAAGTAATGACATTGATAGTCTCTTGAAACAGCTCTACAAACTGCCTTTGAG 943

Qy    841 TCAACGACATGTAGAATCTGATGGCCTTGTCTTTGATCCAGCACAAATCACAAGTCGAAC 900
          |||||| |||||||||||||||||||| | || || ||||| ||||||||||||||||||
Db    944 TCAACGCCATGTAGAATCTGATGGCCTTATTTTCGACCCAGCGCAAATCACAAGTCGAAC 1003

Qy    901 AGCTAGAGGTGTTGCAGTGCCACACGGAGATCATTACCACTTCATCCCTTACTCTCAAAT 960
          || |||||||| || || || || ||  | |||||||||||| ||||||||    |||||
Db   1004 CGCCAGAGGTGTAGCTGTCCCTCATGGTAACCATTACCACTTTATCCCTTATGAACAAAT 1063

Qy    961 GTCTGAATTGGAAGAACGAATCGCTCGTATTATTCCCCTTCGTTATCGTTCAAACCATTG 1020
          |||||||||||||| ||||||| ||||||||||||||||||||||||||||||||||||
Db   1064 GTCTGAATTGGAAAAACGAATTGCTCGTATTATTCCCCTTCGTTATCGTTCAAACCATTG 1123

Qy   1021 GGTACCAGATTCAAGGCCAGAACAACCAAGTCCACAACCGACTCCGGAACCTAGTCCAGG 1080
          |||||||||||||||| |||||| |||||||||||||||||||||| ||||||||||| |
Db   1124 GGTACCAGATTCAAGACCAGAAGAACCAAGTCCACAACCGACTCCAGAACCTAGTCCAAG 1183

Qy   1081 CCCGCAACCTGCACCAAATCTTAAAATAGACTCAA-----------ATTCTTCTTTGGT 1128
          |||||||||||||||||| | ||  ||   |||            ||      |||||
Db   1184 TCCGCAACCTGCACCAAATCCTCAACCAGCTCCAAGCAATCCAATTGATGAGAAATTGGT 1243

Qy   1129 TAGTCAGCTGGTACGAAAAGTTGGGGAAGGATATGTATTCGAAGAAAAGGGCATCTCTCG 1188
          |   |    || |||||||| || || || ||||| || || || || ||  | |||||
Db   1244 CAAAGAAGCTGTTCGAAAAGTAGGCGATGGTTATGTCTTTGAGGAGAATGGAGTTTCTCG 1303

Qy   1189 TTATGTCTTTGCGAAAGATTTACCATCTGAAACTGTTAAAAATCTTGAAAGCAAGTTATC 1248
          |||| ||   || ||  || |  || | ||||| |     |    |||| ||||| |  |
Db   1304 TTATATCCCAGCCAAGAATCTTTCAGCAGAAACAGCAGCAGGCATTGATAGCAAACTGGC 1363

Qy   1249 AAAACAAGAGAGTGTTTCACACACTTTAACTGCTAAAAAAGAAAATGTTGCTCCTCGTGA 1308
          || || || ||| | || || |   ||  ||||| |||    |  |  |  || ||||
Db   1364 CAAGCAGGAAAGTTTATCTCATAAGCTAGGAGCTAAGAAAACTGACCTCCCATCTAGTGA 1423

Qy   1309 CCAAGAATTTTATGATAAAGCATATAATCTGTTAACTGAGGCTCATAAAGCCTTGTTTGN 1368
          | |||||||||  |||| || ||| |  |  || |      |||  |||  ||  |||
```

```
Db   1424 TCGAGAATTTTACAATAAGGCTTATGACTTACTAGCAAGAATTCACCAAGATTTACTTGA 1483

Qy   1369 AAATAAGGGTCGTAATTCTGATTTCCAAGCCTTAGACAAATTATTAGAACGCTTGAATGA 1428
          ||||| ||||| |   |||||| | || || || ||  | || |||| | || ||
Db   1484 TAATAAAGGTCGACAAGTTGATTTTGAGGCTTTGGATAACCTGTTGGAACGACTCAAGGA 1543

Qy   1429 TGAATCGACTAATAAAGAAAAATTGGTAGATGATTTATTGGCATTCCTAGCACCAATTAC 1488
          ||  || | | |||||| || || || |||||| | | || ||| |||| || |||
Db   1544 TGTCTCAAGTGATAAAGTCAAGTTAGTGGATGATATTCTTGCCTTCTTAGCTCCGATTCG 1603

Qy   1489 CCATCCAGAGCGACTTGGCAAACCAAATTCTCAAATTGAGTATACTGAAGACGAAGTTCG 1548
          |||||||| ||  | || ||||||||| | ||||||  || ||||| || ||  |||
Db   1604 TCATCCAGAACGTTTAGGAAAACCAAATGCGCAAATTACCTACACTGATGATGAGATTCA 1663

Qy   1549 TATTGCTCAATTAGCTGATAAGTATACAACGTCAGATGGTTACATTTTTGATGAACATGA 1608
            | ||  | || || |  ||||| |||||   ||| ||||| || ||||||   | |||
Db   1664 AGTAGCCAAGTTGGCAGGCAAGTACACAACAGAAGACGGTTATATCTTTGATCCTCGTGA 1723

Qy   1609 TATAATCAGTGATGAAGGAGATGCATATGTAACGCCTCATATGGGCCATAGTCACTGGAT 1668
          ||||| ||||||||| || ||||| |||||||| || |||||| |||||| ||||||||
Db   1724 TATAACCAGTGATGAGGGGGATGCCTATGTAACTCCACATATGACCCATAGCCACTGGAT 1783

Qy   1669 TGGAAAAGATAGCCTTTCTGATAAGGAAAAAGTTGCAGCTCAAGCCTATACTAAAGAAAA 1728
          |  |||||||||  | |||||    || | ||  ||||| || || ||| ||||||| ||
Db   1784 TAAAAAAGATAGTTTGTCTGAAGCTGAGAGAGCGGCAGCCCAGGCTTATGCTAAAGAGAA 1843

Qy   1729 AGGTATCCTACCTCCATCTCCAGACGCAGATGTTAAAGCAAATCCAACTGGAGATAGTGC 1788
          |||| |    ||||| ||  |||||    | | |  || |||| |    | | |  | ||
Db   1844 AGGTTTGACCCCTCCTTCGACAGACCATCAGGATTCAGGAAATACTGAGGCAAAAGGAGC 1903

Qy   1789 AGCAGCTATTTACAATCGTGTGAAAGGGGAAAAACGAATTCCACTCGTTCGACTTCCATA 1848
          || |||||| ||||| || ||||||| | ||    | ||||| | |||  | || ||
Db   1904 AGAAGCTATCTACAACCGCGTGAAAGCAGCTAAGAAGGTGCCACTTGATCGTATGCCTTA 1963

Qy   1849 TATGGTTGAGCATACAGTTGAGGTTAAAAACGGTAATTTGATTATTCCTCATAAGGATCA 1908
          |   || |  ||| || || || |||||||||| ||| || || |||||| | || ||
Db   1964 CAATCTTCAATATACTGTAGAAGTCAAAAACGGTAGTTTAATCATACCTCATTATGACCA 2023

Qy   1909 TTACCATAATATTAAATTTGCTTGGTTTGATGATCACACATACAAAGCTCCAAATGGCTA 1968
          ||||||||| || ||||||| |||||||| ||    |   ||  | || || || || ||
Db   2024 TTACCATAACATCAAATTTGAGTGGTTTGACGAAGGCCTTTATGAGGCACCTAAGGGGTA 2083

Qy   1969 TACCTTGGAAGATTTGTTTGCGACGATTAAGTACTACGTAGAACACCCTGACGAACGTCC 2028
          |||  | || ||| | || |||||  | |||||||| || ||||| ||  ||||||||||
Db   2084 TACTCTTGAGGATCTTTTGGCGACTGTCAAGTACTATGTCGAACATCCAAACGAACGTCC 2143

Qy   2029 ACATTCTAATGATGGATGGGGCAATGCCAGTGAGCATGTGTTAGGCAAGAAAGACCACAG 2088
          |||||  || |||| |  || || || || || |||||   | | || ||| |
Db   2144 GCATTCAGATAATGGTTTTGGTAACGCTAGCGACCATGTTCAAAGAAACAAAAATGGTCA 2203

Qy   2089 TGAAGATCCAAATAAGAACTTCAAAGCGGATGAAGAG---------------------- 2125
          |  ||| | ||| | |     ||| |    || |||
Db   2204 AGCTGATACCAATCAAACGGAAAAACCAAGCGAGGAGAAACCTCAGACAGAAAAACCTGA 2263

Qy   2126 ------------------------------------------------------------ 2125

Db   2264 GGAAGAAACCCCTCGAGAAGAGAAACCACAAAGCGAGAAACCAGAGTCTCCAAAACCAAC 2323

Qy   2126 -------CCAGTAGAGGAAACACCTGCTGAGCCAGAAGTCCCTCAAGTAGAGACTGAAAA 2178
                |||| ||| ||| |||| |  ||  ||||||  ||||| || ||||||||||
Db   2324 AGAGGAACCAGAAGAAGAATCACCAGAGGAATCAGAAGAACCTCAGGTCGAGACTGAAAA 2383

Qy   2179 AGTAGAAGCCCAACTCAAAGAAGCAGAAGTTTTGCTTGCGAAAGTAACGGATTCTAGTCT 2238
          || ||||   |||| | ||| || |||| ||| ||||  ||| |   |||| | | | |
Db   2384 GGTTGAAGAAAAACTGAGAGAGGCTGAAGATTTACTTGGAAAAATCCAGGATCCAATTAT 2443

Qy   2239 GAAAGCCAATGCAACAGAAACTCTAGCTGGTTTACGAAATAATTTGACTCTTCAAATTAT 2298
          ||  ||||||| | ||| |||||  | || |||  |||||||||     ||   |
Db   2444 CAAGTCCAATGCCAAAGAGACTCTCACAGGATTAAAAAATAATTTACTATTTGGCACCCA 2503

Qy   2299 GGATAACAATAGTATCATGGCAGAAGCAGAAAAATTACTTGCGTTGTTAAAAGGAAGTAA 2358
          ||| ||||||| ||| |||||||||||| |||||| || | || || ||||| |  |||||
Db   2504 GGACAACAATACTATTATGGCAGAAGCTGAAAAACTATTGGCTTTATTAAAGGAGAGTAA 2563
```

RESULT 8
AAA08557
ID AAA08557 standard; DNA; 2478 BP.
XX
AC AAA08557;
XX
DT 19-JUL-2000 (first entry)
XX
DE S. pneumoniae 92 kDa human C3-degrading protein coding sequence.
XX
KW Human C3-degrading protein; 92 kDa; immunostimulatory; vaccine;
KW inhibitor; inflammation; organ rejection; xenotransplantation; ss.
XX
OS Streptococcus pneumoniae.
XX
PN WO200017370-A1.
XX
PD 30-MAR-2000.
XX
PF 24-SEP-1999; 99WO-US22362.
XX
PR 24-SEP-1998; 98US-0101736.
PR 31-MAR-1999; 99US-0283094.
XX
PA (MINU) UNIV MINNESOTA.
PA (AMCY) AMERICAN CYANAMID CO.
XX
PI Hostetter MK, Finkel DJ, Cheng Q, Green BA, Masi AW;
XX
DR WPI; 2000-283594/24.
DR P-PSDB; AAY91939.
XX
PT Isolated polypeptide is used to stimulate immune system and immunize or
PT treat a mammalian subject against Streptococcus pneumoniae infection or
PT colonization
XX
PS Claim 1; Page 55-57; 63pp; English.
XX
CC The present sequence, isolated from Streptococcus pneumoniae, encodes a
CC human C3-degrading protein (see AAY91939) of about 92 kDa. This
CC sequence may encompass a smaller 20 kDa polypeptide coding sequence
CC (AAA08556) also having human C3-degrading activity.
CC The DNA sequences can be used for producing an immune response to
CC Streptococcus pneumoniae in a mammal. Antibodies against the proteins
CC can be used to inhibit S. pneumoniae-mediated C3 degradation.
CC C3-mediated inflammation and rejection in xenotransplantation can be
CC inhibited by expressing the nucleic acid sequences on the surface of an
CC organ of an animal. In particular, the polypeptides are useful for
CC stimulating the immune system and are effective to immunize or treat a
CC mammalian subject against Streptococcus pneumoniae infection or
CC colonization.
XX
SQ Sequence 2478 BP; 837 A; 510 C; 535 G; 596 T; 0 other;

Query Match 42.4%; Score 1011.8; DB 21; Length 2478;
Best Local Similarity 66.2%; Pred. No. 3.8e-244;
Matches 1628; Conservative 0; Mismatches 688; Indels 144; Gaps 6;

```
Qy      1 TTCTTACGAGTTGGGACTGTATCAAGCTAGAACGGTTAAGGAAAA---TAATCGTGTTTC 57
          ||| || ||  | || |   | |||||| |   ||||||| || |   |||||| |||||
Db     60 TTCCTATGAACTTGGTCGTCACCAAGCTGGTCAGGTTAAGAAAGAGTCTAATCGAGTTTC 119

Qy     58 CTATATAGATGGAAAACAAGCGACGCAAAAAACGGAGAATTTGACTCCTGATGAGGTTAG 117
            ||||||||||  | || ||    |||||  | || || ||||| || ||||| || ||
Db    120 TTATATAGATGGTGATCAGGCTGGTCAAAAGGCAGAAAACTTGACACCAGATGAAGTCAG 179

Qy    118 CAAGCGTGAAGGAATCAATGCTGAGCAAATCGTCATCAAGATAACAGACCAAGGCTATGT 177
          ||| | || || ||||| || || ||||||||||||||||| || || ||||| |||||
Db    180 TAAGAGGGAGGGGATCAACGCCGAACAAATCGTCATCAAGATTACGGATCAAGGTTATGT 239

Qy    178 CACTTCACATGGCGACCACTATCATTATTACAATGGTAAGGTTCCTTATGACGCTATCAT 237
```


Db 240 GACCTCTCATGGAGACCATTATCATTACTATAATGGCAAGGTCCCTTATGATGCCATCAT 299
Qy 238 CAGTGAAGAATTACTCATGAAAGATCCAAACTATAAGCTAAAAGATGAGGATATTGTTAA 297
Db 300 CAGTGAAGAGCTCCTCATGAAAGATCCGAATTATCAGTTGAAGGATTCAGACATTGTCAA 359
Qy 298 TGAGGTCAAGGGTGGATATGTTATCAAGGTAGATGGAAAATACTATGTTTACCTTAAGGA 357
Db 360 TGAAATCAAGGGTGGTTATGTTATCAAGGTAGATGGAAAATACTATGTTTACCTTAAGGA 419
Qy 358 TGCTGCCCACGCGGATAACGTCCGTACAAAAGAGGAAATCAATCGACAAAAACAAGAGCA 417
Db 420 TGCAGCTCATGCGGATAATATTCGGACAAAAGAAGAGATTAAACGTCAGAAGCAGGAACA 479
Qy 418 TAGTCAACATCGTGAAGGTGGAACTCCAAGAAACGATGGTGCTGTTGCCTTGGCACGTTC 477
Db 480 CAGTCATAATCACGGGGGTGGTTCT------AACGATCAAGCAGTAGTTGCAGCCAGAGC 533
Qy 478 GCAAGGACGCTATACTACAGATGATGGTTATATCTTTAATGCTTCTGATATCATAGAGGA 537
Db 534 CCAAGGACGCTATACAACGGATGATGGTTATATCTTCAATGCATCTGATATCATTGAGGA 593
Qy 538 TACTGGTGATGCTTATATCGTTCCTCATGGAGATCATTACCATTACATTCCTAAGAATGA 597
Db 594 CACGGGTGATGCTTATATCGTTCCTCACGGCGACCATTACCATTACATTCCTAAGAATGA 653
Qy 598 GTTATCAGCTAGCGAGTTGGCTGCTGCAGAAGCCTTCCTATCTGGTCGAGGAAATCTGTC 657
Db 654 GTTATCAGCTAGCGAGTTAGCTGCTGCAGAAGCCTATTGGAATGG--------------- 698
Qy 658 AAATTCAAGAACCTATCGCCGACAAAATAGCGATAACACTTCAAGAACAAACTGGGTACC 717
Db 699 -GAAGCAGGGATCTCGTCCTTCTTCAAGTTCTAGTTATAATGCAAATCCAGCTCAACCAA 757
Qy 718 TTCTGTAAGCAATCCAGGAACTACAAATACTAACACAAGCAACAACAGCAACACTAACAG 777
Db 758 GATTGTCAGAGAACCACAATCT-------------------GACTGTCACTCCAACTTA 797
Qy 778 TCAAGCAAGTCAAAGTAATGACATTGATAGTCTCTTGAAACAGCTCTACAAACTGCCTTT 837
Db 798 TCATCAAAATCAAGGGGAAAACATTTCAAGCCTTTTACGTGAATTGTATGCTAAACCCTT 857
Qy 838 GAGTCAACGACATGTAGAATCTGATGGCCTTGTCTTTGATCCAGCACAAATCACAAGTCG 897
Db 858 ATCAGAACGCCATGTGGAATCTGATGGCCTTATTTTCGACCCAGCGCAAATCACAAGTCG 917
Qy 898 AACAGCTAGAGGTGTTGCAGTGCCACACGGAGATCATTACCACTTCATCCCTTACTCTCA 957
Db 918 AACCGCCAGAGGTGTAGCTGTCCCTCATGGTAACCATTACCACTTTATCCCTTATGAACA 977
Qy 958 AATGTCTGAATTGGAAGAACGAATCGCTCGTATTATTCCCCTTCGTTATCGTTCAAACCA 1017
Db 978 AATGTCTGAATTGGAAAAACGAATTGCTCGTATTATTCCCCTTCGTTATCGTTCAAACCA 1037
Qy 1018 TTGGGTACCAGATTCAAGGCCAGAACAACCAAGTCCACAACCGACTCCGGAACCTAGTCC 1077
Db 1038 TTGGGTACCAGATTCAAGACCAGAACAACCAAGTCCACAATCGACTCCGGAACCTAGTCC 1097
Qy 1078 AGGCCCGCAACCTGCACCAAATCTTAAAATAGACTCAA-----------ATTCTTCTTT 1125
Db 1098 AAGTCCGCAACCTGCACCAAATCCTCAACCAGCTCCAAGCAATCCAATTGATGAGAAATT 1157
Qy 1126 GGTTAGTCAGCTGGTACGAAAAGTTGGGGAAGGATATGTATTCGAAGAAAAGGGCATCTC 1185
Db 1158 GGTCAAAGAAGCTGTTCGAAAAGTAGGCGATGGTTATGTCTTTGAGGAGAATGGAGTTTC 1217
Qy 1186 TCGTTATGTCTTTGCGAAAGATTTACCATCTGAAACTGTTAAAAATCTTGAAAGCAAGTT 1245
Db 1218 TCGTTATATCCCAGCCAAGGATCTTTCAGCAGAAACAGCAGCAGGCATTGATAGCAAACT 1277
Qy 1246 ATCAAAACAAGAGAGTGTTTCACACACTTTAACTGCTAAAAAAGAAAATGTTGCTCCTCG 1305
Db 1278 GGCCAAGCAGGAAAGTTTATCTCATAAGCTAGGAGCTAAGAAAACTGACCTCCCATCTAG 1337
Qy 1306 TGACCAAGAATTTTATGATAAAGCATATAATCTGTTAACTGAGGCTCATAAAGCCTTGTT 1365
Db 1338 TGATCGAGAATTTTACAATAAGGCTTATGACTTACTAGCAAGAATTCACCAAGATTTACT 1397
Qy 1366 TGNAAATAAGGGTCGTAATTCTGATTTCCAAGCCTTAGACAAATTATTAGAACGCTTGAA 1425
Db 1398 TGATAATAAAGGTCGACAAGTTGATTTTGAGGCTTTGGATAACCTGTTGGAACGACTCAA 1457
Qy 1426 TGATGAATCGACTAATAAAGAAAAATTGGTAGATGATTTATTGGCATTCCTAGCACCAAT 1485
Db 1458 GGATGTCCCAAGTGATAAAGTCAAGTTAGTGGATGATATTCTTGCCTTCTTAGCTCCGAT 1517
Qy 1486 TACCCATCCAGAGCGACTTGGCAAACCAAATTCTCAAATTGAGTATACTGAAGACGAAGT 1545
Db 1518 TCGTCATCCAGAACGTTTAGGAAAACCAAATGCGCAAATTACCTACACTGATGATGAGAT 1577
Qy 1546 TCGTATTGCTCAATTAGCTGATAAGTATACAACGTCAGATGGTTACATTTTTGATGAACA 1605
Db 1578 TCAAGTAGCCAAGTTGGCAGGCAAGTACACAACAGAAGACGGTTATATCTTTGATCCTCG 1637
Qy 1606 TGATATAATCAGTGATGAAGGAGATGCATATGTAACGCCTCATATGGGCCATAGTCACTG 1665
Db 1638 TGATATAACCAGTGATGAGGGGGATGCCTATGTAACTCCACATATGACCCATAGCCACTG 1697
Qy 1666 GATTGGAAAAGATAGCCTTTCTGATAAGGAAAAAGTTGCAGCTCAAGCCTATACTAAAGA 1725
Db 1698 GATTAAAAAAGATAGTTTGTCTGAAGCTGAGAGAGCGGCAGCCCAGGCTTATGCTAAAGA 1757
Qy 1726 AAAAGGTATCCTACCTCCATCTCCAGACGCAGATGTTAAAGCAAATCCAACTGGAGATAG 1785
Db 1758 GAAAGGTTTGACCCCTCCTTCGACAGACCATCAGGATTCAGGAAATACTGAGGCAAAAGG 1817
Qy 1786 TGCAGCAGCTATTTACAATCGTGTGAAAGGGGAAAAACGAATTCCACTCGTTCGACTTCC 1845
Db 1818 AGCAGAAGCTATCTACAACCGCGTGAAAGCAGCTAAGAAGGTGCCACTTGATCGTATGCC 1877
Qy 1846 ATATATGGTTGAGCATACAGTTGAGGTTAAAAACGGTAATTTGATTATTCCTCATAAGGA 1905
Db 1878 TTACAATCTTCAATATACTGTAGAAGTCAAAAACGGTAGTTTAATCATACCTCATTATGA 1937
Qy 1906 TCATTACCATAATATTAAATTTGCTTGGTTTGATGATCACACATACAAAGCTCCAAATGG 1965
Db 1938 CCATTACCATAACATCAAATTTGAGTGGTTTGACGAAGGCCTTTATGAGGCACCTAAGGG 1997
Qy 1966 CTATACCTTGGAAGATTTGTTTGCGACGATTAAGTACTACGTAGAACACCCTGACGAACG 2025
Db 1998 GTATACTCTTGAGGATCTTTTGGCGACTGTCAAGTACTATGTCGAACATCCAAACGAACG 2057
Qy 2026 TCCACATTCTAATGATGGATGGGGCAATGCCAGTGAGCATGTGTTAGGCAAGAAAGACCA 2085
Db 2058 TCCGCATTCAGATAATGGTTTTGGTAACGCTAGCGACCATGTTCAAAGAAACAAAAATGG 2117
Qy 2086 CAGTGAAGATCCAAATAAGAACTTCAAAGCGGATGA---------- 2121
Db 2118 TCAAGCTGATACCAATCAAACGGAAAAACCAAGCGAGGAGAAACCTCAGACAGAAAAACC 2177
Qy 2122 ---------- 2121
Db 2178 TGAGGAAGAAACCCCTCGAGAAGAGAAACCGCAAAGCGAGAAACCAGAGTCTCCAAAACC 2237
Qy 2122 ---AGAGCCAGTAGAGGAAACACCTGCTGAGCCAGAAGTCCCTCAAGTAGAGACTGAAAA 2178
Db 2238 AACAGAGGAACCAGAAGAATCACCAGAGGAATCAGAAGAACCTCAGGTCGAGACTGAAAA 2297
Qy 2179 AGTAGAAGCCCAACTCAAAGAAGCAGAAGTTTTGCTTGCGAAAGTAACGGATTCTAGTCT 2238
Db 2298 GGTTGAAGAAAAACTGAGAGAGGCTGAAGATTTACTTGGAAAAATCCAGGATCCAATTAT 2357
Qy 2239 GAAAGCCAATGCAACAGAAACTCTAGCTGGTTTACGAAATAATTTGACTCTTCAAATTAT 2298
Db 2358 CAAGTCCAATGCCAAAGAGACTCTCACAGGATTAAAAAATAATTTACTATTTGGCACCCA 2417

```
Qy   2299 GGATAACAATAGTATCATGGCAGAAGCAGAAAAATTACTTGCGTTGTTAAAAGGAAGTAA 2358
          ||| |||||||| ||| ||||||||||| |||||| || | || || ||||| |  |||||
Db   2418 GGACAACAATACTATTATGGCAGAAGCTGAAAAACTATTGGCTTTATTAAAGGAGAGTAA 2477


RESULT 9
AAA05417
ID   AAA05417 standard; DNA; 2481 BP.
XX
AC   AAA05417;
XX
DT   24-MAY-2000  (first entry)
XX
DE   Streptococcus pneumoniae nucleotide sequence ID311.
XX
KW   Streptococcus pneumoniae; vaccine; screening; protein antigen;
KW   antibacterial; antiinflammatory; meningitis; infection; diagnosis;
KW   pneumococcal disease; ds.
XX
OS   Streptococcus pneumoniae.
XX
PN   WO200006737-A2.
XX
PD   10-FEB-2000.
XX
PF   27-JUL-1999;   99WO-GB02451.
XX
PR   27-JUL-1998;   98GB-0016337.
PR   19-MAR-1999;   99US-0125164.
XX
PA   (MICR-) MICROBIAL TECHNICS LTD.
XX
PI   Gilbert CFG,  Hansbro PM;
XX
DR   WPI; 2000-195300/17.
DR   P-PSDB; AAY81662.
XX
PT   New Streptococcal protein, useful as a vaccine, for diagnosis of
PT   pneumococcal diseases and for screening agents capable of antagonizing
PT   or inhibiting expression of the protein  -
XX
PS   Claim 2; Page 99; 108pp; English.
XX
CC   AAY81501 to AAY81679 represent specifically claimed protein sequences
CC   isolated from Streptococcus pneumoniae. AAA05407 to AAA05590 represent
CC   specifically claimed nucleotide sequences isolated from S. pneumoniae.
CC   The sequences have antibacterial and antiinflammatory properties.
CC   The protein sequences, and fragments of them, are useful as immunogens
CC   and/or antigens. The nucleotide sequences can be used in vaccines and in
CC   diagnostic assays. The proteins and nucleotides can be useful for the
CC   detection and diagnosis of S. pneumoniae. The protein sequences are also
CC   useful for screening an agent capable of antagonising, inhibiting or
CC   interfering with the function or expression of the proteins in which the
CC   agent is useful for treatment or prophylaxis of S. pneumoniae infection
CC   and meningitis. AAA05591 to AAA05614 represent primers used in the
CC   exemplification of the present invention.
XX
SQ   Sequence 2481 BP; 839 A; 514 C; 538 G; 590 T; 0 other;

  Query Match             42.0%;  Score 1003.8;  DB 21;  Length 2481;
  Best Local Similarity   66.0%;  Pred. No. 3.9e-242;
  Matches 1623;  Conservative    0;  Mismatches  693;  Indels  144;  Gaps    6;

Qy       1 TTCTTACGAGTTGGGACTGTATCAAGCTAGAACGGTTAAGGAAAA---TAATCGTGTTTC 57
             ||| || ||| | || |   | |||||| |   || |||| || |   |||||| ||| |
Db      60 TTCCTATGAGCTTGGTCGTCACCAAGCTGGTCAGGATAAGAAAGAGTCTAATCGAGTTGC 119

Qy      58 CTATATAGATGGAAAACAAGCGACGCAAAAAACGGAGAATTTGACTCCTGATGAGGTTAG 117
           ||||||||||||  | || ||    |||||  | || || ||||| || ||||| || ||
Db     120 TTATATAGATGGTGATCAGGCTGGTCAAAAGGCAGAAAACTTGACACCAGATGAAGTCAG 179

Qy     118 CAAGCGTGAAGGAATCAATGCTGAGCAAATCGTCATCAAGATAACAGACCAAGGCTATGT 177
                ||| | || || ||||| || || |||||||||||||||||| || || ||||| |||||
Db     180 TAAGAGGGAGGGGATCAACGCCGAACAAATCGTCATCAAGATTACGGATCAAGGTTATGT 239

Qy     178 CACTTCACATGGCGACCACTATCATTATTACAATGGTAAGGTTCCTTATGACGCTATCAT 237
           || || ||||| ||||| |||||||| || ||||| ||||| |||||||| || |||||
Db     240 GACCTCTCATGGAGACCATTATCATTACTATAATGGCAAGGTCCCTTATGATGCCATCAT 299

Qy     238 CAGTGAAGAATTACTCATGAAAGATCCAAACTATAAGCTAAAAGATGAGGATATTGTTAA 297
           |||||||||  | ||||||||||||||| || ||| || | || |||   || ||||| ||
Db     300 CAGTGAAGAGCTCCTCATGAAAGATCCGAATTATCAGTTGAAGGATTCAGACATTGTCAA 359

Qy     298 TGAGGTCAAGGGTGGATATGTTATCAAGGTAGATGGAAAATACTATGTTTACCTTAAGGA 357
           |||  |||||||||| ||||| |||||||||| ||||||||||||||||||||||||||
Db     360 TGAAATCAAGGGTGGTTATGTCATCAAGGTAGACGGAAAATACTATGTTTACCTTAAGGA 419

Qy     358 TGCTGCCCACGCGGATAACGTCCGTACAAAAGAGGAAATCAATCGACAAAAACAAGAGCA 417
           ||| || || |||||||  | || |||||||| || || || || || || || || | |
Db     420 TGCAGCTCATGCGGATAATATTCGGACAAAAGAAGAGATTAAACGTCAGAAGCAGGAACG 479

Qy     418 TAGTCAACATCGTGAAGGTGGAACTCCAAGAAACGATGGTGCTGTTGCCTTGGCACGTTC 477
           |||||  |||  |     ||| ||      ||||||  ||| || ||    ||  |  |
Db     480 CAGTCATAATCACGGGTCAGGAGCT------AACGATCATGCAGTAGCTGCAGCCAGAGC 533

Qy     478 GCAAGGACGCTATACTACAGATGATGGTTATATCTTTAATGCTTCTGATATCATAGAGGA 537
           |||||||||||||| || |||||||| |||||||| ||||| |||||||||| |||||
Db     534 CCAAGGACGCTATACAACGGATGATGGGTATATCTTCAATGCATCTGATATCATTGAGGA 593

Qy     538 TACTGGTGATGCTTATATCGTTCCTCATGGAGATCATTACCATTACATTCCTAAGAATGA 597
           || ||||||||||||||||||||||| || || ||||||||||||||||||||||||||
Db     594 CACGGGTGATGCTTATATCGTTCCTCACGGCGACCATTACCATTACATTCCTAAGAATGA 653

Qy     598 GTTATCAGCTAGCGAGTTGGCTGCTGCAGAAGCCTTCCTATCTGGTCGAGGAAATCTGTC 657
           ||||||||||||||||| ||||||||||||||        |||
Db     654 GTTATCAGCTAGCGAGTTAGCTGCTGCAGAAGCCTATTGGAATGG--------------- 698

Qy     658 AAATTCAAGAACCTATCGCCGACAAAATAGCGATAACACTTCAAGAACAAACTGGGTACC 717
             |  || | | ||    |      ||   | |       |  | | | | ||
Db     699 -GAAGCAGGGATCTCGTCCTTCTTCAAGTTCTAGTTATAATGCAAATCCAGCTCAACCAA 757

Qy     718 TTCTGTAAGCAATCCAGGAACTACAAATACTAACACAAGCAACAACAGCAACACTAACAG 777
              ||| ||  | |||  | ||                      |    ||   | |
Db     758 GATTGTCAGAGAACCACAATCT-------------------GACTGTCACTCCAACTTA 797

Qy     778 TCAAGCAAGTCAAAGTAATGACATTGATAGTCTCTTGAAACAGCTCTACAAACTGCCTTT 837
           |||   || |||| |  |  |||||   || || ||      |  | ||       || ||
Db     798 TCATCAAAATCAAGGGGAAAACATTTCAAGCCTTTTACGTGAATTGTATGCTAAACCCTT 857

Qy     838 GAGTCAACGACATGTAGAATCTGATGGCCTTGTCTTTGATCCAGCACAAATCACAAGTCG 897
               |||| ||||| |||||||||||||| | || || ||||| ||||||||||||||
Db     858 ATCAGAACGCCATGTGGAATCTGATGGCCTTATTTTCGACCCAGCGCAAATCACAAGTCG 917

Qy     898 AACAGCTAGAGGTGTTGCAGTGCCACACGGAGATCATTACCACTTCATCCCTTACTCTCA 957
           ||| || |||||||| || || || || ||  | |||||||||| ||||||||    ||
Db     918 AACCGCCAGAGGTGTAGCTGTCCCTCATGGTAACCATTACCACTTTATCCCTTATGAACA 977

Qy     958 AATGTCTGAATTGGAAGAACGAATCGCTCGTATTATTCCCCTTCGTTATCGTTCAAACCA 1017
           |||||||||||||||| ||||||| |||||||||||||||||||||||||||||||||||
Db     978 AATGTCTGAATTGGAAAAACGAATTGCTCGTATTATTCCCCTTCGTTATCGTTCAAACCA 1037

Qy    1018 TTGGGTACCAGATTCAAGGCCAGAACAACCAAGTCCACAACCGACTCCGGAACCTAGTCC 1077
           |||||||||||||||||| ||||||||||||||||||| ||||||||||||||||||||
Db    1038 TTGGGTACCAGATTCAAGACCAGAACAACCAAGTCCACAATCGACTCCGGAACCTAGTCC 1097

Qy    1078 AGGCCCGCAACCTGCACCAAATCTTAAAATAGACTCAA-----------ATTCTTCTTT 1125
           | | |||||||||||||||||| | ||  ||  |||           ||      ||
Db    1098 AAGTCCGCAACCTGCACCAAATCCTCAACCAGCTCCAAGCAATCCAATTGATGAGAAATT 1157

Qy    1126 GGTTAGTCAGCTGGTACGAAAAGTTGGGGAAGGATATGTATTCGAAGAAAAGGGCATCTC 1185
           ||| |   |    || |||||||| || || || ||||| || || || || ||  | ||
Db    1158 GGTCAAAGAAGCTGTTCGAAAAGTAGGCGATGGTTATGTCTTTGAGGAGAATGGAGTTTC 1217

Qy    1186 TCGTTATGTCTTTGCGAAAGATTTACCATCTGAAACTGTTAAAAATCTTGAAAGCAAGTT 1245
           |||||||  ||   || || ||| |  || | ||||| |    |    |||| |||||  |
```

```
Db   1218 TCGTTATATCCCAGCCAAGGATCTTTCAGCAGAAACAGCAGCAGGCATTGATAGCAAACT 1277

Qy   1246 ATCAAAACAAGAGAGTGTTTCACACACTTTAACTGCTAAAAAAGAAAATGTTGCTCCTCG 1305
          | || || || ||| | || || |   ||   ||||| |||     |  |  |  || |
Db   1278 GGCCAAGCAGGAAAGTTTATCTCATAAGCTAGGAGCTAAGAAAACTGACCTCCCATCTAG 1337

Qy   1306 TGACCAAGAATTTTATGATAAAGCATATAATCTGTTAACTGAGGCTCATAAAGCCTTGTT 1365
          ||| | |||||||||  |||| || ||| |   |  || |       |||  |||  ||  |
Db   1338 TGATCGAGAATTTTACAATAAGGCTTATGACTTACTAGCAAGAATTCACCAAGATTTACT 1397

Qy   1366 TGNAAATAAGGGTCGTAATTCTGATTTCCAAGCCTTAGACAAATTATTAGAACGCTTGAA 1425
          ||  ||||| |||||  |   ||||||  | || || || ||  | || |||||  | ||
Db   1398 TGATAATAAAGGTCGACAAGTTGATTTTGAGGCTTTGGATAACCTGTTGGAACGACTCAA 1457

Qy   1426 TGATGAATCGACTAATAAAGAAAAATTGGTAGATGATTTATTGGCATTCCTAGCACCAAT 1485
          ||||   | | | ||||||  || || || ||||||  |   | || ||| |||| || ||
Db   1458 GGATGTCCCAAGTGATAAAGTCAAGTTAGTGGATGATATTCTTGCCTTCTTAGCTCCGAT 1517

Qy   1486 TACCCATCCAGAGCGACTTGGCAAACCAAATTCTCAAATTGAGTATACTGAAGACGAAGT 1545
          |  |||||||| ||  | || |||||||||| | ||||||   || |||| || ||  |
Db   1518 TCGTCATCCAGAACGTTTAGGAAAACCAAATGCGCAAATTACCTACACTGATGATGAGAT 1577

Qy   1546 TCGTATTGCTCAATTAGCTGATAAGTATACAACGTCAGATGGTTACATTTTTGATGAACA 1605
          ||   | ||  | || || |  ||||| |||||   ||| ||||| || ||||||   |
Db   1578 TCAAGTAGCCAAGTTGGCAGGCAAGTACACAACAGAAGACGGTTATATCTTTGATCCTCG 1637

Qy   1606 TGATATAATCAGTGATGAAGGAGATGCATATGTAACGCCTCATATGGGCCATAGTCACTG 1665
          |||||||| ||||||||| || ||||| |||||||| || ||||||  ||||||| ||||
Db   1638 TGATATAACCAGTGATGAGGGGGATGCCTATGTAACTCCACATATGACCCATAGCCACTG 1697

Qy   1666 GATTGGAAAAGATAGCCTTTCTGATAAGGAAAAAGTTGCAGCTCAAGCCTATACTAAAGA 1725
          ||||  |||||||||  | |||||    || | ||  ||||| || || ||| |||||||
Db   1698 GATTAAAAAAGATAGTTTGTCTGAAGCTGAGAGAGCGGCAGCCCAGGCTTATGCTAAAGA 1757

Qy   1726 AAAAGGTATCCTACCTCCATCTCCAGACGCAGATGTTAAAGCAAATCCAACTGGAGATAG 1785
          ||||||  |    ||||| ||  |||||    | | |  || |||| |    | | |  |
Db   1758 GAAAGGTTTGACCCCTCCTTCGACAGACCATCAGGATTCAGGAAATACTGAGGCAAAAGG 1817

Qy   1786 TGCAGCAGCTATTTACAATCGTGTGAAAGGGGAAAAACGAATTCCACTCGTTCGACTTCC 1845
          ||||  |||||  |||||  || |||||||  |  ||      | |||||  | |||  | ||
Db   1818 AGCAGAAGCTATCTACAACCGCGTGAAAGCAGCTAAGAAGGTGCCACTTGATCGTATGCC 1877

Qy   1846 ATATATGGTTGAGCATACAGTTGAGGTTAAAAACGGTAATTTGATTATTCCTCATAAGGA 1905
          ||  |   ||  |  |||| || || || ||||||||||  ||| || ||  |||||  | ||
Db   1878 TTACAATCTTCAATATACTGTAGAAGTCAAAAACGGTAGTTTAATCATACCTCATTATGA 1937

Qy   1906 TCATTACCATAATATTAAATTTGCTTGGTTTGATGATCACACATACAAAGCTCCAAATGG 1965
          ||||||||||  || |||||||  |||||||| ||    |   ||  | || || || ||
Db   1938 CCATTACCATAACATCAAATTTGAGTGGTTTGACGAAGGCCTTTATGAGGCACCTAAGGG 1997

Qy   1966 CTATACCTTGGAAGATTTGTTTGCGACGATTAAGTACTACGTAGAACACCCTGACGAACG 2025
          |||||  | || ||| | || |||||  | ||||||||  || |||||  ||  |||||||
Db   1998 GTATACTCTTGAGGATCTTTTGGCGACTGTCAAGTACTATGTCGAACATCCAAACGAACG 2057

Qy   2026 TCCACATTCTAATGATGGATGGGGCAATGCCAGTGAGCATGTGTTAGGCAAGAAAGACCA 2085
          ||| |||||  || |||| |  || || || || || |||||   | | || ||| |
Db   2058 TCCGCATTCAGATAATGGTTTTGGTAACGCTAGCGACCATGTTCAAAGAAACAAAAATGG 2117

Qy   2086 CAGTGAAGATCCAAATAAGAACTTCAAAGCGGATGA---------------------- 2121
               |  ||| | ||| | |     ||| |    ||
Db   2118 TCAAGCTGATACCAATCAAACGGAAAAACCAAGCGAGGAGAAACCTCAGACAGAAAAACC 2177

Qy   2122 ------------------------------------------------------------ 2121

Db   2178 TGAGGAAGAAACCCCTCGAGAAGAGAAACCGCAAAGCGAGAAACCAGAGTCTCCAAAACC 2237

Qy   2122 ---AGAGCCAGTAGAGGAAACACCTGCTGAGCCAGAAGTCCCTCAAGTAGAGACTGAAAA 2178
             ||||  |  ||| ||| |||| |  ||  |||||  ||||| || ||||||||||
Db   2238 AACAGAGGAACCAGAAGAATCACCAGAGGAATCAGAAGAACCTCAGGTCGAGACTGAAAA 2297

Qy   2179 AGTAGAAGCCCAACTCAAAGAAGCAGAAGTTTTGCTTGCGAAAGTAACGGATTCTAGTCT 2238
          ||  ||||   |||| | ||| || |||| ||| ||||  |||  |   |||| | | | |
Db   2298 GGTTGAAGAAAAACTGAGAGAGGCTGAAGATTTACTTGGAAAAATCCAGGATCCAATTAT 2357

Qy   2239 GAAAGCCAATGCAACAGAAACTCTAGCTGGTTTACGAAATAATTTGACTCTTCAAATTAT 2298
          ||  ||||||| | ||| |||||  | || |||  |||||||||     ||   |
Db   2358 CAAGTCCAATGCCAAAGAGACTCTCACAGGATTAAAAAATAATTTACTATTTGGCACCCA 2417

Qy   2299 GGATAACAATAGTATCATGGCAGAAGCAGAAAAATTACTTGCGTTGTTAAAAGGAAGTAA 2358
          ||| ||||||| ||| ||||||||||| |||||| || | || || ||||| |  |||||
Db   2418 GGACAACAATACTATTATGGCAGAAGCTGAAAAACTATTGGCTTTATTAAAGGAGAGTAA 2477
```

RESULT 10
AAA47605
ID AAA47605 standard; DNA; 2531 BP.
XX
AC AAA47605;
XX
DT 20-OCT-2000 (first entry)
XX
DE Recombinant variant of Sp36 gene (Sp36B) of S. pneumoniae.
XX
KW Streptococcus pneumoniae; infection; vaccine; coiled coil region;
KW histidine triad residue; Sp36; antibody; otitis media;
KW nasopharyngeal infection; bronchial infection; bronchitis; sepsis;
KW meningitis; lobar pneumonia; ds.
XX
OS Streptococcus pneumoniae.
XX
FH Key Location/Qualifiers
FT CDS 1..2531
FT /*tag= a
FT /product= Sp36B polypeptide
XX
PN WO200037105-A2.
XX
PD 29-JUN-2000.
XX
PF 21-DEC-1999; 99WO-US30390.
XX
PR 21-DEC-1998; 98US-0113048.
XX
PA (MEDI-) MEDIMMUNE INC.
XX
PI Johnson LS, Koenig S, Adamou JE;
XX
DR WPI; 2000-452129/39.
DR P-PSDB; AAB01469.
XX
PT Vaccine useful for prophylaxis and treatment of pnemococcal infections
PT such as otitis media, nasopharyngeal and bronchial infections,
PT comprises Streptococcus pneumoniae proteins
XX
PS Disclosure; Page 69-70; 70pp; English.
XX
CC Although a number of proteins have been suggested as being involved
CC in the pathogenicity of Streptococcus pneumoniae, there still remains
CC a need to identify polypeptides having epitopes in common from
CC various strains of S. pneumoniae in order to utilise such
CC polypeptides in vaccines to protect against a wide variety of
CC S. pneumoniae. New vaccine compositions are described which comprise a
CC Streptococcus pneumoniae polypeptide (or fragments) of 80 - 680 amino
CC acids in length that comprise at least one histidine triad residue
CC (HxxHxH) or a coiled-coil region, or an antibody directed against
CC these features. The vaccine is useful in protecting against infection
CC by Streptococcus pneumoniae. The vaccine composition comprising
CC antibodies to is useful for passive immunization for treating
CC Pneumococcal infections which includes otitis media, nasopharyngeal
CC and bronchial infections.
XX
SQ Sequence 2531 BP; 861 A; 508 C; 550 G; 609 T; 3 other;

Query Match 41.5%; Score 991; DB 21; Length 2531;
Best Local Similarity 65.7%; Pred. No. 6.6e-239;
Matches 1609; Conservative 1; Mismatches 697; Indels 141; Gaps 6;

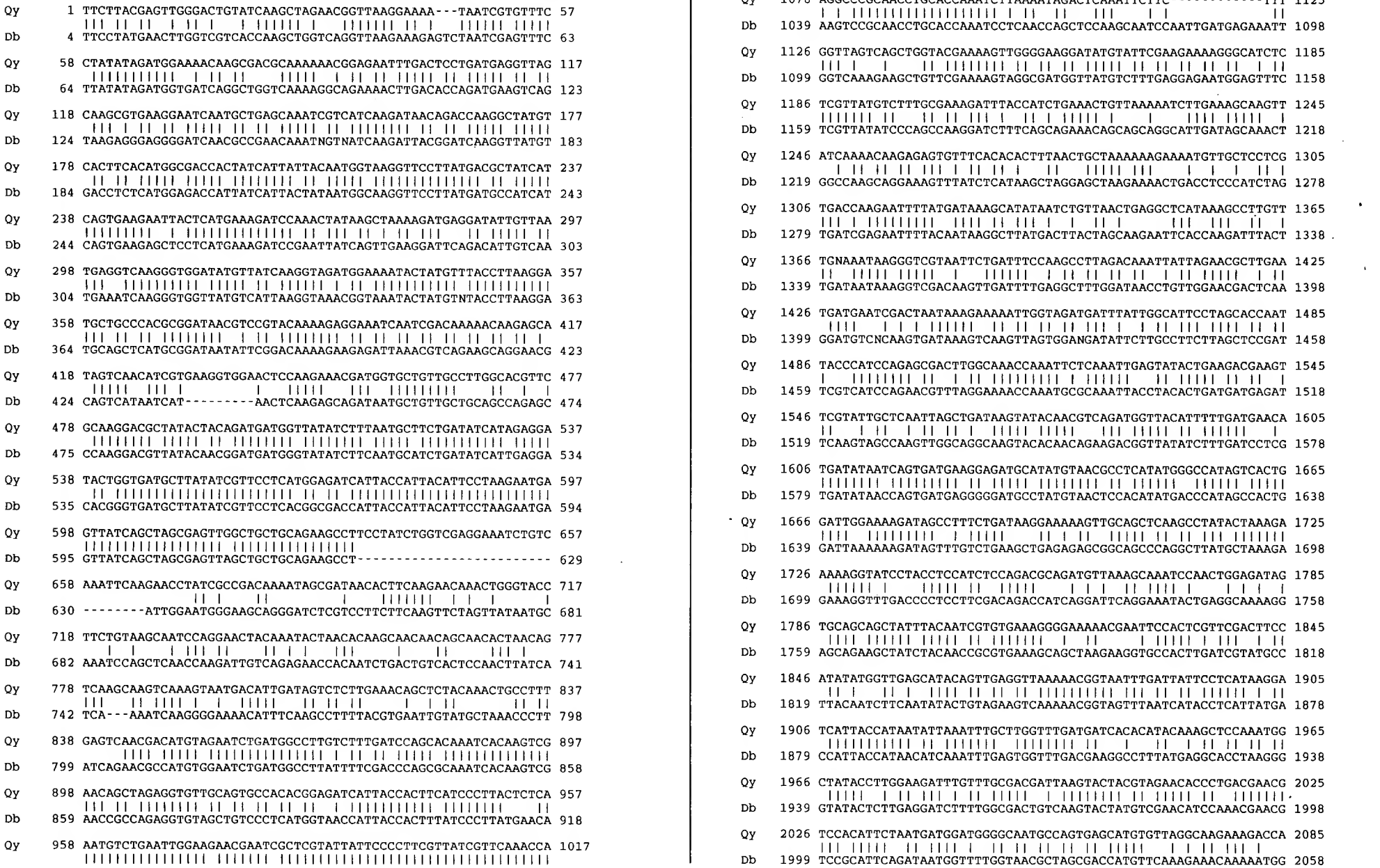
```
Db   2169 CCCTCGAGAAGAGAAACCGCAAAGCGAGAAACCAGAGTCTCCAAAACCAACAGAGGAACC 2228

Qy   2131 AGAGGAAACACCTGCTGAGCCAGAAGTCCCTCAAGTAGAGACTGAAAAAGTAGAAGCCCA 2190

Db   2229 AGAAGAATCACCAGAGGAATCAGAAGAACCTCAGGTCGAGACTGAAAAGGTTGAAGAAAA 2288

Qy   2191 ACTCAAAGAAGCAGAAGTTTTGCTTGCGAAAGTAACGGATTCTAGTCTGAAAGCCAATGC 2250

Db   2289 ACTGAGAGAGGCTGAAGATTTACTTGGAAAAATCCAGGATCCAATTATCAAGTCCAATGC 2348

Qy   2251 AACAGAAACTCTAGCTGGTTTACGAAATAATTTGACTCTTCAAATTATGGATAACAATAG 2310

Db   2349 CAAAGAGACTCTCACAGGATTAAAAAATAATTTACTATTTGGCACCCAGGACAACAATAC 2408

Qy   2311 TATCATGGCAGAAGCAGAAAAATTACTTGCGTTGTTAAAAGGAAGTAA 2358

Db   2409 TATTATGGCAGAAGCTGAAAAACTATTGGCTTTATTAAAGGAGAGTAA 2456


RESULT 11
AAA47602
ID   AAA47602 standard; DNA; 2531 BP.
XX
AC   AAA47602;
XX
DT   20-OCT-2000  (first entry)
XX
DE   Recombinant variant of Sp36 gene (Sp36D) of S. pneumoniae.
XX
KW   Streptococcus pneumoniae; infection; vaccine; coiled coil region;
KW   histidine triad residue; Sp36; antibody; otitis media;
KW   nasopharyngeal infection; bronchial infection; bronchitis; sepsis;
KW   meningitis; lobar pneumonia; ds.
XX
OS   Streptococcus pneumoniae.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..2517
FT                   /*tag= a
FT                   /product= Sp36D polypeptide
XX
PN   WO200037105-A2.
XX
PD   29-JUN-2000.
XX
PF   21-DEC-1999;   99WO-US30390.
XX
PR   21-DEC-1998;   98US-0113048.
XX
PA   (MEDI-) MEDIMMUNE INC.
XX
PI   Johnson LS,  Koenig S,  Adamou JE;
XX
DR   WPI; 2000-452129/39.
DR   P-PSDB; AAB01466.
XX
PT   Vaccine useful for prophylaxis and treatment of pnemococcal infections
PT   such as otitis media, nasopharyngeal and bronchial infections,
PT   comprises Streptococcus pneumoniae proteins
XX
PS   Disclosure; Page 57-58; 70pp; English.
XX
CC   Although a number of proteins have been suggested as being involved
CC   in the pathogenicity of Streptococcus pneumoniae, there still remains
CC   a need to identify polypeptides having epitopes in common from
CC   various strains of S. pneumoniae in order to utilise such
CC   polypeptides in vaccines to protect against a wide variety of
CC   S. pneumoniae. New vaccine compositions are described which comprise a
CC   Streptococcus pneumoniae polypeptide (or fragments) of 80 - 680 amino
CC   acids in length that comprise at least one histidine triad residue
CC   (HxxHxH) or a coiled-coil region, or an antibody directed against
CC   these features. The vaccine is useful in protecting against infection
CC   by Streptococcus pneumoniae. The vaccine composition comprising
CC   antibodies to is useful for passive immunization for treating
CC   Pneumococcal infections which includes otitis media, nasopharyngeal
CC   and bronchial infections.
XX
SQ   Sequence 2531 BP; 836 A; 511 C; 547 G; 637 T; 0 other;

  Query Match             41.4%;  Score 990;  DB 21;  Length 2531;
  Best Local Similarity   68.1%;  Pred. No. 1.2e-238;
  Matches 1478;  Conservative    0;  Mismatches  636;  Indels   57;  Gaps    5;

Qy      1 TTCTTACGAGTTGGGACTGTATCAAGCTAGAACGGTTAAGGAAAA---TAATCGTGTTTC 57

Db     60 TTCCTATGAACTTGGTCGTCACCAAGCTGGTCAGGTTAAGAAAGAGTCTAATCGAGTTTC 119

Qy     58 CTATATAGATGGAAAACAAGCGACGCAAAAAACGGAGAATTTGACTCCTGATGAGGTTAG 117

Db    120 TTATATAGATGGTGATCAGGCTGGTCAAAAGGCAGAAAACTTGACACCAGATGAAGTCAG 179

Qy    118 CAAGCGTGAAGGAATCAATGCTGAGCAAATCGTCATCAAGATAACAGACCAAGGCTATGT 177

Db    180 TAAGAGGGAGGGGATCAACGCCGAACAAATCGTCATCAAGATTACGGATCAAGGTTATGT 239

Qy    178 CACTTCACATGGCGACCACTATCATTATTACAATGGTAAGGTTCCTTATGACGCTATCAT 237

Db    240 GACCTCTCATGGAGACCATTATCATTACTATAATGGCAAGGTCCCTTATGATGCCATCAT 299

Qy    238 CAGTGAAGAATTACTCATGAAAGATCCAAACTATAAGCTAAAAGATGAGGATATTGTTAA 297

Db    300 CAGTGAAGAGCTCCTCATGAAAGATCCGAATTATCAGTTGAAGGATTCAGACATTGTCAA 359

Qy    298 TGAGGTCAAGGGTGGATATGTTATCAAGGTAGATGGAAAATACTATGTTTACCTTAAGGA 357

Db    360 TGAAATCAAGGGTGGTTATGTTATCAAGGTAGATGGAAAATACTATGTTTACCTTAAGGA 419

Qy    358 TGCTGCCCACGCGGATAACGTCCGTACAAAAGAGGAAATCAATCGACAAAAACAAGAGCA 417

Db    420 TGCAGCTCATGCGGATAATATTCGGACAAAAGAAGAGATTAAACGTCAGAAGCAGGAACA 479

Qy    418 TAGTCAACATCGTGAAGGTGGAACTCCAAGAAACGATGGTGCTGTTGCCTTGGCACGTTC 477

Db    480 CAGTCATAATCACGGGGGTGGTTCT------AACGATCAAGCAGTAGTTGCAGCCAGAGC 533

Qy    478 GCAAGGACGCTATACTACAGATGATGGTTATATCTTTAATGCTTCTGATATCATAGAGGA 537

Db    534 CCAAGGACGCTATACAACGGATGATGGTTATATCTTCAATGCATCTGATATCATTGAGGA 593

Qy    538 TACTGGTGATGCTTATATCGTTCCTCATGGAGATCATTACCATTACATTCCTAAGAATGA 597

Db    594 CACGGGTGATGCTTATATCGTTCCTCACGGCGACCATTACCATTACATTCCTAAGAATGA 653

Qy    598 GTTATCAGCTAGCGAGTTGGCTGCTGCAGAAGCCTTCCTATCTGGTCGAGGAAATCTGTC 657

Db    654 GTTATCAGCTAGCGAGTTAGCTGCTGCAGAAGCCT------------------------- 688

Qy    658 AAATTCAAGAACCTATCGCCGACAAAATAGCGATAACACTTCAAGAACAAACTGGGTACC 717

Db    689 --------ATTGGAATGGGAAGCAGGGATCTCGTCCTTCTTCAAGTTCTAGTTATAATGC 740

Qy    718 TTCTGTAAGCAATCCAGGAACTACAAATACTAACACAAGCAACAACAGCAACACTAACAG 777

Db    741 AAATCCAGCTCAACCAAGATTGTCAGAGAACCACAATCTGACTGTCACTCCAACTTATCA 800

Qy    778 TCAAGCAAGTCAAAGTAATGACATTGATAGTCTCTTGAAACAGCTCTACAAACTGCCTTT 837

Db    801 TCA---AAATCAAGGGGAAAACATTTCAAGCCTTTTACGTGAATTGTATGCTAAACCCTT 857

Qy    838 GAGTCAACGACATGTAGAATCTGATGGCCTTGTCTTTGATCCAGCACAAATCACAAGTCG 897

Db    858 ATCAGAACGCCATGTGGAATCTGATGGCCTTATTTTCGACCCAGCGCAAATCACAAGTCG 917

Qy    898 AACAGCTAGAGGTGTTGCAGTGCCACACGGAGATCATTACCACTTCATCCCTTACTCTCA 957
```

```
Db    918 AACCGCCAGAGGTGTAGCTGTCCCTCATGGTAACCATTACCACTTTATCCCTTATGAACA 977

Qy    958 AATGTCTGAATTGGAAGAACGAATCGCTCGTATTATTCCCCTTCGTTATCGTTCAAACCA 1017
          ||||||||||||||||| ||||||| ||||||||||||||||||||||||||||||||||
Db    978 AATGTCTGAATTGGAAAAACGAATTGCTCGTATTATTCCCCTTCGTTATCGTTCAAACCA 1037

Qy   1018 TTGGGTACCAGATTCAAGGCCAGAACAACCAAGTCCACAACCGACTCCGGAACCTAGTCC 1077
          ||||||||||||||||| ||||||||||||||||||| ||||||||||||||||||||
Db   1038 TTGGGTACCAGATTCAAGACCAGAACAACCAAGTCCACAATCGACTCCGGAACCTAGTCC 1097

Qy   1078 AGGCCCGCAACCTGCACCAAATCTTAAAATAGACTCAAATTCTTC----------TTT 1125
          | | ||||||||||||||||||| | ||  ||   |||    | |           ||
Db   1098 AAGTCCGCAACCTGCACCAAATCCTCAACCAGCTCCAAGCAATCCAATTGATGAGAAATT 1157

Qy   1126 GGTTAGTCAGCTGGTACGAAAAGTTGGGGAAGGATATGTATTCGAAGAAAAGGGCATCTC 1185
          ||| |   |    || |||||||| || || || ||||| || || || || ||  | ||
Db   1158 GGTCAAAGAAGCTGTTCGAAAAGTAGGCGATGGTTATGTCTTTGAGGAGAATGGAGTTTC 1217

Qy   1186 TCGTTATGTCTTTGCGAAAGATTTACCATCTGAAACTGTTAAAAATCTTGAAAGCAAGTT 1245
          ||||||| ||   || || ||| |  || | ||||| |    |    |||| |||||  |
Db   1218 TCGTTATATCCCAGCCAAGGATCTTTCAGCAGAAACAGCAGCAGGCATTGATAGCAAACT 1277

Qy   1246 ATCAAAACAAGAGAGTGTTTCACACACTTTAACTGCTAAAAAAGAAAATGTTGCTCCTCG 1305
            | || || || ||| | || || |   ||   ||||| |||    |  |  |  || |
Db   1278 GGCCAAGCAGGAAAGTTTATCTCATAAGCTAGGAGCTAAGAAAACTGACCTCCCATCTAG 1337

Qy   1306 TGACCAAGAATTTTATGATAAAGCATATAATCTGTTAACTGAGGCTCATAAAGCCTTGTT 1365
          ||| | |||||||||  |||| || ||| |  |  || |      |||  |||  ||  |
Db   1338 TGATCGAGAATTTTACAATAAGGCTTATGACTTACTAGCAAGAATTCACCAAGATTTACT 1397

Qy   1366 TGNAAATAAGGGTCGTAATTCTGATTTCCAAGCCTTAGACAAATTATTAGAACGCTTGAA 1425
          ||  ||||| |||||  |   ||||||  | || || || ||  | || |||||  | ||
Db   1398 TGATAATAAAGGTCGACAAGTTGATTTTGAGGCTTTGGATAACCTGTTGGAACGACTCAA 1457

Qy   1426 TGATGAATCGACTAATAAAGAAAAATTGGTAGATGATTTATTGGCATTCCTAGCACCAAT 1485
            ||||   | | | |||||  || || || ||||| |  | || ||| |||| || ||
Db   1458 GGATGTCCCAAGTGATAAAGTCAAGTTAGTGGATGATATTCTTGCCTTCTTAGCTCCGAT 1517

Qy   1486 TACCCATCCAGAGCGACTTGGCAAACCAAATTCTCAAATTGAGTATACTGAAGACGAAGT 1545
          |  |||||||| ||  | || ||||||||| | ||||||   || ||||| || ||  |
Db   1518 TCGTCATCCAGAACGTTTAGGAAAACCAAATGCGCAAATTACCTACACTGATGATGAGAT 1577

Qy   1546 TCGTATTGCTCAATTAGCTGATAAGTATACAACGTCAGATGGTTACATTTTTGATGAACA 1605
          ||   | ||  | || || |  ||||| |||||   ||| ||||| || ||||||   |
Db   1578 TCAAGTAGCCAAGTTGGCAGGCAAGTACACAACAGAAGACGGTTATATCTTTGATCCTCG 1637

Qy   1606 TGATATAATCAGTGATGAAGGAGATGCATATGTAACGCCTCATATGGGCCATAGTCACTG 1665
          |||||||| ||||||||| || ||||| |||||||| || ||||||  ||||||  |||||
Db   1638 TGATATAACCAGTGATGAGGGGGATGCCTATGTAACTCCACATATGACCCATAGCCACTG 1697

Qy   1666 GATTGGAAAAGATAGCCTTTCTGATAAGGAAAAAGTTGCAGCTCAAGCCTATACTAAAGA 1725
          ||||  |||||||||  | |||||    || | ||  ||||| || || ||| |||||||
Db   1698 GATTAAAAAAGATAGTTTGTCTGAAGCTGAGAGAGCGGCAGCCCAGGCTTATGCTAAAGA 1757

Qy   1726 AAAAGGTATCCTACCTCCATCTCCAGACGCAGATGTTAAAGCAAATCCAACTGGAGATAG 1785
          ||||||  |    ||||| ||  |||||    | | |  || |||| |    | | |  |
Db   1758 GAAAGGTTTGACCCCTCCTTCGACAGACCATCAGGATTCAGGAAATACTGAGGCAAAAGG 1817

Qy   1786 TGCAGCAGCTATTTACAATCGTGTGAAAGGGGAAAAACGAATTCCACTCGTTCGACTTCC 1845
          |||| |||||| ||||| || |||||||  |  ||     | ||||| | |||  | ||
Db   1818 AGCAGAAGCTATCTACAACCGCGTGAAAGCAGCTAAGAAGGTGCCACTTGATCGTATGCC 1877

Qy   1846 ATATATGGTTGAGCATACAGTTGAGGTTAAAAACGGTAATTTGATTATTCCTCATAAGGA 1905
            || |   || |  |||| || || || ||||||||| ||| || || |||||| | ||
Db   1878 TTACAATCTTCAATATACTGTAGAAGTCAAAAACGGTAGTTTAATCATACCTCATTATGA 1937

Qy   1906 TCATTACCATAATATTAAATTTGCTTGGTTTGATGATCACACATACAAAGCTCCAAATGG 1965
          ||||||||||| || ||||||| |||||||| ||   |   ||  | || || || ||
Db   1938 CCATTACCATAACATCAAATTTGAGTGGTTTGACGAAGGCCTTTATGAGGCACCTAAGGG 1997

Qy   1966 CTATACCTTGGAAGATTTGTTTGCGACGATTAAGTACTACGTAGAACACCCTGACGAACG 2025
          |||||  | || ||| | || |||||  | |||||||| || ||||| ||  |||||||
Db   1998 GTATACTCTTGAGGATCTTTTGGCGACTGTCAAGTACTATGTCGAACATCCAAACGAACG 2057

Qy   2026 TCCACATTCTAATGATGGATGGGGCAATGCCAGTGAGCATGTGTTAGGCAAGAAAGACCA 2085
          ||| |||||  || |||| |  || || || || || |||||       || || |   |
Db   2058 TCCGCATTCAGATAATGGTTTTGGTAACGCTAGCGACCATGTTCGTAAAAATAAGGTAGA 2117

Qy   2086 CAGTGAAGATCCAAATAAGAACTTCAAAGCGGATGAAGAGCCAGTAGAGGAAACACCTGC 2145
          |   ||   |  | | |  |    || |   |||| ||   |   |||  ||| |   |
Db   2118 CCAAGACAGTAAACCTGATGAAGATAAGGAACATGATGAAGTAAGTGAGCCAACTCACCC 2177

Qy   2146 TGAGCCAGAAG 2156
          |||  | || |
Db   2178 TGAATCTGATG 2188
```

RESULT 12
AAV27356
ID AAV27356 standard; DNA; 2290 BP.
XX
AC AAV27356;
XX
DT 02-OCT-1998 (first entry)
XX
DE Streptococcus pneumoniae SP0042 nucleotide.
XX
KW Streptococcus pneumoniae; antigen; vaccine; infection; diagnosis;
KW detection; pneumonia; otitis media; meningitis; ss.
XX
OS Streptococcus pneumoniae.
XX
FH Key   Location/Qualifiers
FT CDS   2..2290
FT   /*tag= a
FT   /product= "SP0042"
FT   /transl_except= (pos:152..154,aa:Xaa)
FT   /transl_except= (pos:1406..1408,aa:Xaa)
FT   /transl_except= (pos:1430..1432,aa:Xaa)
FT   /note= "no stop codon given; Xaa is unspecified"
XX
PN WO9818930-A2.
XX
PD 07-MAY-1998.
XX
PF 30-OCT-1997; 97WO-US19422.
XX
PR 31-OCT-1996; 96US-0029960.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Choi GH, Hromockyj A, Johnson LS, Kunsch CA;
XX
DR WPI; 1998-272224/24.
DR P-PSDB; AAW55095.
XX
PT Nucleic acid encoding antigenic peptide(s) from Streptococcus
PT pneumoniae - or their epitope-containing fragments, useful in
PT protective or therapeutic vaccines, and for diagnosis
XX
PS Claim 1; Page 61-62; 118pp; English.
XX
CC The present sequence encodes a protein from Streptococcus pneumoniae.
CC The nucleic acid sequence encoding the Streptococcus pneumoniae protein
CC can be useful in vaccines for inducing protective antibodies against
CC Streptococcus pneumoniae, for treatment or prevention of infection e.g.
CC pneumonia, otitis media or meningitis. Probes based on the nucleic acid
CC are used to detect Streptococcus infection (by usual hybridisation or
CC amplification methods), also for isolating Streptococcus genes or their
CC allelic variants. The protein can be used similarly to detect specific
CC antibodies in standard immunoassays, especially for diagnosing or
CC monitoring infections. Antibodies which bind the protein are used to
CC detect corresponding antigens, to purify the protein and for passive
CC immunisation (optionally coupled to a toxin). Vaccines are administered,
CC e.g. by injection, orally or through the skin, typically at 0.01-1000
CC (especially 10-300) mu g/ml per dose.

XX
SQ Sequence 2290 BP; 766 A; 474 C; 498 G; 547 T; 5 other;

Query Match 41.3%; Score 987.6; DB 19; Length 2290;
Best Local Similarity 67.7%; Pred. No. 4.5e-238;
Matches 1481; Conservative 0; Mismatches 645; Indels 60; Gaps 5;



```
Qy   2086 CAGTGAAGATCCAAATAAGAACTTCAAAGCGGATGAAGAGCCAGTAGAGGAAACACCTGC 2145
Db   2059 TCAAGCTGATACCAATCAAACGGAAAAACCAAGCGAGGAGAAACCTCAGACAGAAAAACC 2118

Qy   2146 TGAGCCAGAAGTCCCTCAAGTAGAGA 2171
Db   2119 TGAGGAAGAAACCCCTCGAGAAGAGA 2144
```

RESULT 13
ABQ84824
ID ABQ84824 standard; DNA; 2290 BP.
XX
AC ABQ84824;
XX
DT 04-SEP-2002 (first entry)
XX
DE S. pneumoniae SP042 nucleotide sequence SEQ ID NO:65.
XX
KW Streptococcus pneumoniae; epitope; vaccine; antigenic protein;
KW antibacterial; Streptococcal infection; detection; gene; ds.
XX
OS Streptococcus pneumoniae.
XX
PN US2002061545-A1.
XX
PD 23-MAY-2002.
XX
PF 22-JAN-2001; 2001US-0765272.
XX
PR 30-OCT-1997; 97US-0961083.
XX
PA (CHOI/) CHOI G H.
PA (KUNS/) KUNSCH C A.
PA (BARA/) BARASH S C.
PA (DILL/) DILLON P J.
PA (DOUG/) DOUGHERTY B.
PA (FANN/) FANNON M R.
PA (ROSE/) ROSEN C A.
XX
PI Choi GH, Kunsch CA, Barash SC, Dillon PJ, Dougherty B, Fannon MR;
PI Rosen CA;
XX
DR WPI; 2002-479261/51.
DR P-PSDB; ABP54589.
XX
PT New Streptococcus pneumoniae antigens, useful for detecting
PT Streptococcus and for preventing or attenuating disease caused by
PT Streptococcus infection -
XX
PS Claim 1; Page 28-29; 70pp; English.
XX
CC ABQ84792 to ABQ84904 represents nucleic acids which encode the
CC Streptococcus pneumoniae antigens given in ABP54557 to ABP54669.
CC The S. pneumoniae antigens have antibacterial activity and can be
CC used in vaccines. The S. pneumoniae antigens can also be used to
CC prevent or attenuate a Streptococcal infection in an animal. The
CC polynucleotides encoding the S. pneumoniae antigens can be used to
CC detect Streptococcus nucleic acids. ABQ84905 to ABQ85130 represent
CC primers used in the cloning of S. pneumoniae ORFs (open reading frames)
CC which are used in an example from the present invention.
XX
SQ Sequence 2290 BP; 766 A; 474 C; 498 G; 547 T; 5 other;

```
  Query Match            41.3%;  Score 987.6;  DB 24;  Length 2290;
  Best Local Similarity  67.7%;  Pred. No. 4.5e-238;
  Matches 1481;  Conservative    0;  Mismatches  645;  Indels   60;  Gaps    5;

Qy       1 TTCTTACGAGTTGGGACTGTATCAAGCTAGAACGGTTAAGGAAAA---TAATCGTGTTTC 57
Db       4 TTCCTATGAACTTGGTCGTCACCAAGCTGGTCAGGTTAAGAAAGAGTCTAATCGAGTTTC 63

Qy      58 CTATATAGATGGAAAACAAGCGACGCAAAAAACGGAGAATTTGACTCCTGATGAGGTTAG 117
Db      64 TTATATAGATGGTGATCAGGCTGGTCAAAAGGCAGAAAACTTGACACCAGATGAAGTCAG 123

Qy     118 CAAGCGTGAAGGAATCAATGCTGAGCAAATCGTCATCAAGATAACAGACCAAGGCTATGT 177
Db     124 TAAGAGGGAGGGGATCAACGCCGAACAAATNGTNATCAAGATTACGGATCAAGGTTATGT 183

Qy     178 CACTTCACATGGCGACCACTATCATTATTACAATGGTAAGGTTCCTTATGACGCTATCAT 237
Db     184 GACCTCTCATGGAGACCATTATCATTACTATAATGGCAAGGTTCCTTATGATGCCATCAT 243

Qy     238 CAGTGAAGAATTACTCATGAAAGATCCAAACTATAAGCTAAAAGATGAGGATATTGTTAA 297
Db     244 CAGTGAAGAGCTCCTCATGAAAGATCCGAATTATCAGTTGAAGGATTCAGACATTGTCAA 303

Qy     298 TGAGGTCAAGGGTGGATATGTTATCAAGGTAGATGGAAAATACTATGTTTACCTTAAGGA 357
Db     304 TGAAATCAAGGGTGGTTATGTCATTAAGGTAAACGGTAAATACTATGTNTACCTTAAGGA 363

Qy     358 TGCTGCCCACGCGGATAACGTCCGTACAAAAGAGGAAATCAATCGACAAAAACAAGAGCA 417
Db     364 TGCAGCTCATGCGGATAATATTCGGACAAAAGAAGAGATTAAACGTCAGAAGCAGGAACG 423

Qy     418 TAGTCAACATCGTGAAGGTGGAACTCCAAGAAACGATGGTGCTGTTGCCTTGGCACGTTC 477
Db     424 CAGTCATAATCAT---------AACTCAAGAGCAGATAATGCTGTTGCTGCAGCCAGAGC 474

Qy     478 GCAAGGACGCTATACTACAGATGATGGTTATATCTTTAATGCTTCTGATATCATAGAGGA 537
Db     475 CCAAGGACGTTATACAACGGATGATGGGTATATCTTCAATGCATCTGATATCATTGAGGA 534

Qy     538 TACTGGTGATGCTTATATCGTTCCTCATGGAGATCATTACCATTACATTCCTAAGAATGA 597
Db     535 CACGGGTGATGCTTATATCGTTCCTCACGGCGACCATTACCATTACATTCCTAAGAATGA 594

Qy     598 GTTATCAGCTAGCGAGTTGGCTGCTGCAGAAGCCTTCCTATCTGGTCGAGGAAATCTGTC 657
Db     595 GTTATCAGCTAGCGAGTTAGCTGCTGCAGAAGCCT------------------------- 629

Qy     658 AAATTCAAGAACCTATCGCCGACAAAATAGCGATAACACTTCAAGAACAAACTGGGTACC 717
Db     630 --------ATTGGAATGGGAAGCAGGGATCTCGTCCTTCTTCAAGTTCTAGTTATAATGC 681

Qy     718 TTCTGTAAGCAATCCAGGAACTACAAATACTAACACAAGCAACAACAGCAACACTAACAG 777
Db     682 AAATCCAGCTCAACCAAGATTGTCAGAGAACCACAATCTGACTGTCACTCCAACTTATCA 741

Qy     778 TCAAGCAAGTCAAAGTAATGACATTGATAGTCTCTTGAAACAGCTCTACAAACTGCCTTT 837
Db     742 TCA---AAATCAAGGGGAAAACATTTCAAGCCTTTTACGTGAATTGTATGCTAAACCCTT 798

Qy     838 GAGTCAACGACATGTAGAATCTGATGGCCTTGTCTTTGATCCAGCACAAATCACAAGTCG 897
Db     799 ATCAGAACGCCATGTGGAATCTGATGGCCTTATTTTCGACCCAGCGCAAATCACAAGTCG 858

Qy     898 AACAGCTAGAGGTGTTGCAGTGCCACACGGAGATCATTACCACTTCATCCCTTACTCTCA 957
Db     859 AACCGCCAGAGGTGTAGCTGTCCCTCATGGTAACCATTACCACTTTATCCCTTATGAACA 918

Qy     958 AATGTCTGAATTGGAAGAACGAATCGCTCGTATTATTCCCCTTCGTTATCGTTCAAACCA 1017
Db     919 AATGTCTGAATTGGAAAAACGAATTGCTCGTATTATTCCCCTTCGTTATCGTTCAAACCA 978

Qy    1018 TTGGGTACCAGATTCAAGGCCAGAACAACCAAGTCCACAACCGACTCCGGAACCTAGTCC 1077
Db     979 TTGGGTACCAGATTCAAGACCAGAACAACCAAGTCCACAATCGACTCCGGAACCTAGTCC 1038

Qy    1078 AGGCCCGCAACCTGCACCAAATCTTAAAATAGACTCAAATTCTTC------------TTT 1125
Db    1039 AAGTCCGCAACCTGCACCAAATCCTCAACCAGCTCCAAGCAATCCAATTGATGAGAAATT 1098

Qy    1126 GGTTAGTCAGCTGGTACGAAAAGTTGGGGAAGGATATGTATTCGAAGAAAAGGGCATCTC 1185
```

```
Db  1099 GGTCAAAGAAGCTGTTCGAAAAGTAGGCGATGGTTATGTCTTTGAGGAGAATGGAGTTTC 1158

Qy  1186 TCGTTATGTCTTTGCGAAAGATTTACCATCTGAAACTGTTAAAAATCTTGAAAGCAAGTT 1245

Db  1159 TCGTTATATCCCAGCCAAGGATCTTTCAGCAGAAACAGCAGCAGGCATTGATAGCAAACT 1218

Qy  1246 ATCAAAACAAGAGAGTGTTTCACACACTTTAACTGCTAAAAAAGAAAATGTTGCTCCTCG 1305

Db  1219 GGCCAAGCAGGAAAGTTTATCTCATAAGCTAGGAGCTAAGAAAACTGACCTCCCATCTAG 1278

Qy  1306 TGACCAAGAATTTTATGATAAAGCATATAATCTGTTAACTGAGGCTCATAAAGCCTTGTT 1365

Db  1279 TGATCGAGAATTTTACAATAAGGCTTATGACTTACTAGCAAGAATTCACCAAGATTTACT 1338

Qy  1366 TGNAAATAAGGGTCGTAATTCTGATTTCCAAGCCTTAGACAAATTATTAGAACGCTTGAA 1425

Db  1339 TGATAATAAAGGTCGACAAGTTGATTTTGAGGCTTTGGATAACCTGTTGGAACGACTCAA 1398

Qy  1426 TGATGAATCGACTAATAAAGAAAAATTGGTAGATGATTTATTGGCATTCCTAGCACCAAT 1485

Db  1399 GGATGTCNCAAGTGATAAAGTCAAGTTAGTGGANGATATTCTTGCCTTCTTAGCTCCGAT 1458

Qy  1486 TACCCATCCAGAGCGACTTGGCAAACCAAATTCTCAAATTGAGTATACTGAAGACGAAGT 1545

Db  1459 TCGTCATCCAGAACGTTTAGGAAAACCAAATGCGCAAATTACCTACACTGATGATGAGAT 1518

Qy  1546 TCGTATTGCTCAATTAGCTGATAAGTATACAACGTCAGATGGTTACATTTTTGATGAACA 1605

Db  1519 TCAAGTAGCCAAGTTGGCAGGCAAGTACACAACAGAAGACGGTTATATCTTTGATCCTCG 1578

Qy  1606 TGATATAATCAGTGATGAAGGAGATGCATATGTAACGCCTCATATGGGCCATAGTCACTG 1665

Db  1579 TGATATAACCAGTGATGAGGGGGATGCCTATGTAACTCCACATATGACCCATAGCCACTG 1638

Qy  1666 GATTGGAAAAGATAGCCTTTCTGATAAGGAAAAAGTTGCAGCTCAAGCCTATACTAAAGA 1725

Db  1639 GATTAAAAAAGATAGTTTGTCTGAAGCTGAGAGAGCGGCAGCCCAGGCTTATGCTAAAGA 1698

Qy  1726 AAAAGGTATCCTACCTCCATCTCCAGACGCAGATGTTAAAGCAAATCCAACTGGAGATAG 1785

Db  1699 GAAAGGTTTGACCCCTCCTTCGACAGACCATCAGGATTCAGGAAATACTGAGGCAAAAGG 1758

Qy  1786 TGCAGCAGCTATTTACAATCGTGTGAAAGGGGAAAAACGAATTCCACTCGTTCGACTTCC 1845

Db  1759 AGCAGAAGCTATCTACAACCGCGTGAAAGCAGCTAAGAAGGTGCCACTTGATCGTATGCC 1818

Qy  1846 ATATATGGTTGAGCATACAGTTGAGGTTAAAAACGGTAATTTGATTATTCCTCATAAGGA 1905

Db  1819 TTACAATCTTCAATATACTGTAGAAGTCAAAAACGGTAGTTTAATCATACCTCATTATGA 1878

Qy  1906 TCATTACCATAATATTAAATTTGCTTGGTTTGATGATCACACATACAAAGCTCCAAATGG 1965

Db  1879 CCATTACCATAACATCAAATTTGAGTGGTTTGACGAAGGCCTTTATGAGGCACCTAAGGG 1938

Qy  1966 CTATACCTTGGAAGATTTGTTTGCGACGATTAAGTACTACGTAGAACACCCTGACGAACG 2025

Db  1939 GTATACTCTTGAGGATCTTTTGGCGACTGTCAAGTACTATGTCGAACATCCAAACGAACG 1998

Qy  2026 TCCACATTCTAATGATGGATGGGGCAATGCCAGTGAGCATGTGTTAGGCAAGAAAGACCA 2085

Db  1999 TCCGCATTCAGATAATGGTTTTGGTAACGCTAGCGACCATGTTCAAAGAAACAAAAATGG 2058

Qy  2086 CAGTGAAGATCCAAATAAGAACTTCAAAGCGGATGAAGAGCCAGTAGAGGAAACACCTGC 2145

Db  2059 TCAAGCTGATACCAATCAAACGGAAAAACCAAGCGAGGAGAAACCTCAGACAGAAAAACC 2118

Qy  2146 TGAGCCAGAAGTCCCTCAAGTAGAGA 2171

Db  2119 TGAGGAAGAAACCCCTCGAGAAGAGA 2144

RESULT 14
```

```
AAA65737
ID   AAA65737 standard; DNA; 2639 BP.
XX
AC   AAA65737;
XX
DT   21-NOV-2000  (first entry)
XX
DE   Streptococcus pneumoniae BVH-11-2 gene SEQ ID NO:13.
XX
KW   Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;
KW   prophylaxis; therapy; infection; diagnosis; meningitis; bacteraemia;
KW   otitis media; pneumonia; immunisation; bactericidal; ds.
XX
OS   Streptococcus pneumoniae.
XX
PN   WO200039299-A2.
XX
PD   06-JUL-2000.
XX
PF   20-DEC-1999;   99WO-CA01218.
XX
PR   23-DEC-1998;   98US-0113800.
XX
PA   (BIOC-) BIOCHEM PHARMA INC.
XX
PI   Hamel J,  Brodeur BR,  Pineau I,  Martin D,  Rioux C,  Charland N;
XX
DR   WPI; 2000-452397/39.
DR   P-PSDB; AAB12720.
XX
PT   Streptococcal antigens useful for vaccinating against e.g. meningitis,
PT   otitis media, bacteremia and/or pneumonia -
XX
PS   Example 6; Fig 16; 106pp; English.
XX
CC   The present invention describes nucleic acids (I) encoding protein
CC   antigens (II) from Streptococcus pneumoniae. The protein antigens
CC   have bactericidal activity. The nucleic acids, encoding the protein
CC   antigens, may be used for the recombinant production of the proteins
CC   they encode. The protein antigens may then be used as vaccines for the
CC   prevention and treatment of Streptococcal infections in mammals
CC   (especially humans) which result in, e.g. meningitis, otitis media,
CC   bacteraemia and/or pneumonia. The present sequence encodes the
CC   S. pneumoniae BVH-11-2 protein antigen.
XX
SQ   Sequence 2639 BP; 889 A; 518 C; 567 G; 665 T; 0 other;

  Query Match             41.1%;  Score 980.8;  DB 21;  Length 2639;
  Best Local Similarity   67.8%;  Pred. No. 2.5e-236;
  Matches 1473;  Conservative    0;  Mismatches  638;  Indels   60;  Gaps    5;

Qy       1 TTCTTACGAGTTGGGACTGTATCAAGCTAGAACGGTTAAGGAAAA---TAATCGTGTTTC 57

Db     173 TTCCTATGAACTTGGTCGTCACCAAGCTGGTCAGGTTAAGAAAGAGTCTAATCGAGTTTC 232

Qy      58 CTATATAGATGGAAAACAAGCGACGCAAAAAACGGAGAATTTGACTCCTGATGAGGTTAG 117

Db     233 TTATATAGATGGTGATCAGGCTGGTCAAAAGGCAGAAAATTTGACACCAGATGAAGTCAG 292

Qy     118 CAAGCGTGAAGGAATCAATGCTGAGCAAATCGTCATCAAGATAACAGACCAAGGCTATGT 177

Db     293 TAAGAGAGAGGGGATCAACGCCGAACAAATTGTTATCAAGATTACGGATCAAGGTTATGT 352

Qy     178 CACTTCACATGGCGACCACTATCATTATTACAATGGTAAGGTTCCTTATGACGCTATCAT 237

Db     353 GACCTCTCATGGAGACCATTATCATTACTATAATGGCAAGGTTCCTTATGATGCCATCAT 412

Qy     238 CAGTGAAGAATTACTCATGAAAGATCCAAACTATAAGCTAAAAGATGAGGATATTGTTAA 297

Db     413 CAGTGAAGAACTTCTCATGAAAGATCCGAATTATCAGTTGAAGGATTCAGACATTGTCAA 472

Qy     298 TGAGGTCAAGGGTGGATATGTTATCAAGGTAGATGGAAAATACTATGTTTACCTTAAGGA 357
```

```
Db    473 TGAAATCAAGGGTGGCTATGTGATTAAGGTAGACGGAAAATACTATGTTTACCTTAAAGA 532

Qy    358 TGCTGCCCACGCGGATAACGTCCGTACAAAAGAGGAAATCAATCGACAAAAACAAGAGCA 417
          ||| ||||| ||||| || | || |||||||| || || || || || || || || ||
Db    533 TGCGGCCCATGCGGACAATATTCGGACAAAAGAAGAGATTAAACGTCAGAAGCAGGAACA 592

Qy    418 TAGTCAACATCGTGAAGGTGGAACTCCAAGAAACGATGGTGCTGTTGCCTTGGCACGTTC 477
          |||||  ||| |         |   |||||   |||  |||||||||     ||  |  |
Db    593 CAGTCATAATCAT---------AACTCAAGAGCAGATAATGCTGTTGCTGCAGCCAGAGC 643

Qy    478 GCAAGGACGCTATACTACAGATGATGGTTATATCTTTAATGCTTCTGATATCATAGAGGA 537
          ||||||||| |||||  || |||||||| ||||||||  ||||| ||||||||||| ||||||
Db    644 CCAAGGACGTTATACAACGGATGATGGGTATATCTTCAATGCATCTGATATCATTGAGGA 703

Qy    538 TACTGGTGATGCTTATATCGTTCCTCATGGAGATCATTACCATTACATTCCTAAGAATGA 597
          || ||||||||||||||||||||||| || || ||||||||||||||||||||||||||||
Db    704 CACGGGTGATGCTTATATCGTTCCTCACGGCGACCATTACCATTACATTCCTAAGAATGA 763

Qy    598 GTTATCAGCTAGCGAGTTGGCTGCTGCAGAAGCCTTCCTATCTGGTCGAGGAAATCTGTC 657
          ||||||||||||||||| |||||||||||||||
Db    764 GTTATCAGCTAGCGAGTTAGCTGCTGCAGAAGCCT------------------------ 798

Qy    658 AAATTCAAGAACCTATCGCCGACAAAATAGCGATAACACTTCAAGAACAAACTGGGTACC 717
                    || |    ||          |    |||||||  | |  |       |
Db    799 --------ATTGGAATGGGAAGCAGGGATCTCGTCCTTCTTCAAGTTCTAGTTATAATGC 850

Qy    718 TTCTGTAAGCAATCCAGGAACTACAAATACTAACACAAGCAACAACAGCAACACTAACAG 777
            |  |    | ||| ||    || | |   |||      |    ||      ||| |
Db    851 AAATCCAGTTCAACCAAGATTGTCAGAGAACCACAATCTGACTGTCACTCCAACTTATCA 910

Qy    778 TCAAGCAAGTCAAAGTAATGACATTGATAGTCTCTTGAAACAGCTCTACAAACTGCCTTT 837
          |||   || |||| |   |  |||||   || || ||     |  ||      || ||
Db    911 TCA---AAATCAAGGGGAAAACATTTCAAGCCTTTTACGTGAATTGTATGCTAAACCCTT 967

Qy    838 GAGTCAACGACATGTAGAATCTGATGGCCTTGTCTTTGATCCAGCACAAATCACAAGTCG 897
              |||| |||||||||||||||||||||| | || || ||||| ||||||||||||||
Db    968 ATCAGAACGCCATGTAGAATCTGATGGCCTTATTTTCGACCCAGCGCAAATCACAAGTCG 1027

Qy    898 AACAGCTAGAGGTGTTGCAGTGCCACACGGAGATCATTACCACTTCATCCCTTACTCTCA 957
          ||| || |||||||| || || || || ||   | ||||||||||| ||||||||     ||
Db   1028 AACCGCCAGAGGTGTAGCTGTCCCTCATGGTAACCATTACCACTTTATCCCTTATGAACA 1087

Qy    958 AATGTCTGAATTGGAAGAACGAATCGCTCGTATTATTCCCCTTCGTTATCGTTCAAACCA 1017
          |||||||||||||||| ||||||| ||||||||||||||||||||||||||||||||||||
Db   1088 AATGTCTGAATTGGAAAAACGAATTGCTCGTATTATTCCCCTTCGTTATCGTTCAAACCA 1147

Qy   1018 TTGGGTACCAGATTCAAGGCCAGAACAACCAAGTCCACAACCGACTCCGGAACCTAGTCC 1077
          |||||||||||||||||| ||||||||||||||||||| ||||||||||||||||||||
Db   1148 TTGGGTACCAGATTCAAGACCAGAACAACCAAGTCCACAATCGACTCCGGAACCTAGTCC 1207

Qy   1078 AGGCCCGCAACCTGCACCAAATCTTAAAATAGACTCAAATTCTTC------------TTT 1125
          | | | |||||||||||||||||| | ||  ||   |||     | |             ||
Db   1208 AAGTCTGCAACCTGCACCAAATCCTCAACCAGCTCCAAGCAATCCAATTGATGAGAAATT 1267

Qy   1126 GGTTAGTCAGCTGGTACGAAAAGTTGGGGAAGGATATGTATTCGAAGAAAAGGGCATCTC 1185
          |||  |   |    || |||||||| || || || ||||| || || || || ||  | ||
Db   1268 GGTCAAAGAAGCTGTTCGAAAAGTAGGCGATGGTTATGTCTTTGAGGAGAATGGAGTTTC 1327

Qy   1186 TCGTTATGTCTTTGCGAAAGATTTACCATCTGAAACTGTTAAAAATCTTGAAAGCAAGTT 1245
          |||||||  ||   || || ||| |  || | ||||| |     |    |||| |||||  |
Db   1328 TCGTTATATCCCAGCCAAGGATCTTTCAGCAGAAACAGCAGCAGGCATTGATAGCAAACT 1387

Qy   1246 ATCAAAACAAGAGAGTGTTTCACACACTTTAACTGCTAAAAAAGAAAATGTTGCTCCTCG 1305
            | || || || ||| | || || |   ||   ||||| |||    |  |  |  || |
Db   1388 GGCCAAGCAGGAAAGTTTATCTCATAAGCTAGGAGCTAAGAAAACTGACCTCCCATCTAG 1447

Qy   1306 TGACCAAGAATTTTATGATAAAGCATATAATCTGTTAACTGAGGCTCATAAAGCCTTGTT 1365
          ||| | |||||||||  |||| || ||| |  |  || |      |||  |||  ||  |
Db   1448 TGATCGAGAATTTTACAATAAGGCTTATGACTTACTAGCAAGAATTCACCAAGATTTACT 1507

Qy   1366 TGNAAATAAGGGTCGTAATTCTGATTTCCAAGCCTTAGACAAATTATTAGAACGCTTGAA 1425
          ||  ||||| |||||  |   ||||||  | |  || || ||  | || |||||  | ||
Db   1508 TGATAATAAAGGTCGACAAGTTGATTTTGAGGTTTTGGATAACCTGTTGGAACGACTCAA 1567
```

```
Qy   1426 TGATGAATCGACTAATAAAGAAAAATTGGTAGATGATTTATTGGCATTCCTAGCACCAAT 1485
          ||||  || | | ||||||  || || || |||||| |  | || ||| |||| || ||
Db   1568 GGATGTCTCAAGTGATAAAGTCAAGTTAGTGGATGATATTCTTGCCTTCTTAGCTCCGAT 1627

Qy   1486 TACCCATCCAGAGCGACTTGGCAAACCAAATTCTCAAATTGAGTATACTGAAGACGAAGT 1545
          |   |||||||| ||  | || |||||||| | ||||||   || ||||| || ||  |
Db   1628 TCGTCATCCAGAACGTTTAGGAAAACCAAATGCGCAAATTACCTACACTGATGATGAGAT 1687

Qy   1546 TCGTATTGCTCAATTAGCTGATAAGTATACAACGTCAGATGGTTACATTTTTGATGAACA 1605
          ||   | ||  | || || |  ||||| |||||   ||| ||||| || ||||||   |
Db   1688 TCAAGTAGCCAAGTTGGCAGGCAAGTACACAACAGAAGACGGTTATATCTTTGATCCTCG 1747

Qy   1606 TGATATAATCAGTGATGAAGGAGATGCATATGTAACGCCTCATATGGGCCATAGTCACTG 1665
          |||||||| ||||||||| || ||||| |||||||| || ||||||  ||||||| |||||
Db   1748 TGATATAACCAGTGATGAGGGGGATGCCTATGTAACTCCACATATGACCCATAGCCACTG 1807

Qy   1666 GATTGGAAAAGATAGCCTTTCTGATAAGGAAAAAGTTGCAGCTCAAGCCTATACTAAAGA 1725
          ||||  |||||||||  | |||||     || | ||  ||||| || || ||| |||||||
Db   1808 GATTAAAAAAGATAGTTTGTCTGAAGCTGAGAGAGCGGCAGCCCAGGCTTATGCTAAAGA 1867

Qy   1726 AAAAGGTATCCTACCTCCATCTCCAGACGCAGATGTTAAAGCAAATCCAACTGGAGATAG 1785
          ||||||  |    ||||| ||  |||||     | | |  || |||| |    | | |  |
Db   1868 GAAAGGTTTGACCCCTCCTTCGACAGACCACCAGGATTCAGGAAATACTGAGGCAAAAGG 1927

Qy   1786 TGCAGCAGCTATTTACAATCGTGTGAAAGGGGAAAAACGAATTCCACTCGTTCGACTTCC 1845
          ||||  |||||| ||||| || |||||||  |  ||      | |||||  | |||  | ||
Db   1928 AGCAGAAGCTATCTACAACCGCGTGAAAGCAGCTAAGAAGGTGCCACTTGATCGTATGCC 1987

Qy   1846 ATATATGGTTGAGCATACAGTTGAGGTTAAAAACGGTAATTTGATTATTCCTCATAAGGA 1905
          || |   || |  |||| || || || ||||||||||  ||| || || ||||||  | ||
Db   1988 TTACAATCTTCAATATACTGTAGAAGTCAAAAACGGTAGTTTAATCATACCTCATTATGA 2047

Qy   1906 TCATTACCATAATATTAAATTTGCTTGGTTTGATGATCACACATACAAAGCTCCAAATGG 1965
          ||||||||||| || |||||||  |||||||| ||    |   ||  | || || || ||
Db   2048 CCATTACCATAACATCAAATTTGAGTGGTTTGACGAAGGCCTTTATGAGGCACCTAAGGG 2107

Qy   1966 CTATACCTTGGAAGATTTGTTTGCGACGATTAAGTACTACGTAGAACACCCTGACGAACG 2025
          ||||   | || ||| | || |||||  | |||||||| || ||||| ||  |||||||
Db   2108 GTATAGTCTTGAGGATCTTTTGGCGACTGTCAAGTACTATGTCGAACATCCAAACGAACG 2167

Qy   2026 TCCACATTCTAATGATGGATGGGGCAATGCCAGTGAGCATGTGTTAGGCAAGAAAGACCA 2085
          ||| |||||  || |||| |  || || || ||||| |||||        || || |   |
Db   2168 TCCGCATTCAGATAATGGTTTTGGTAACGCTAGTGACCATGTTCGTAAAAATAAGGCAGA 2227

Qy   2086 CAGTGAAGATCCAAATAAGAACTTCAAAGCGGATGAAGAGCCAGTAGAGGAAACACCTGC 2145
          |   ||   |  |  | |  |    || |   |||| ||   |   |||  ||| |   |
Db   2228 CCAAGATAGTAAACCTGATGAAGATAAGGAACATGATGAAGTAAGTGAGCCAACTCACCC 2287

Qy   2146 TGAGCCAGAAG 2156
          |||  | || |
Db   2288 TGAATCTGATG 2298
```

```
RESULT 15
ABK15104
ID   ABK15104 standard; DNA; 2639 BP.
XX
AC   ABK15104;
XX
DT   08-MAY-2002  (first entry)
XX
DE   DNA encoding Streptococcus pneumoniae BVH-11-2.
XX
KW   BVH-3; BVH-11; vaccine; meningitis; otitis media; bacteraemia;
KW   pneumonia; streptococcal bacterial infection; gene; ds; BVH-11-2.
XX
OS   Streptococcus pneumoniae.
XX
FH   Key             Location/Qualifiers
FT   CDS             114..2630
FT                   /*tag= a
FT                   /product= "BVH-11-2"
```

```
FT                   /note= "The gene is flanked by sequences from the
FT                          vector SP64, no information on which is
FT                          given in the specification"
XX
PN   WO200198334-A2.
XX
PD   27-DEC-2001.
XX
PF   19-JUN-2001; 2001WO-CA00908.
XX
PR   20-JUN-2000; 2000US-212683P.
XX
PA   (SHIR-) SHIRE BIOCHEM INC.
XX
PI   Hamel J,  Ouellet C,  Charland N,  Martin D,  Brodeur B;
XX
DR   WPI; 2002-122272/16.
DR   P-PSDB; AAU75934.
XX
PT   New Streptococcus pneumoniae BVH-3 and BVH-11 variant and
PT   epitope-bearing polypeptides, useful as vaccine components for treating
PT   or preventing streptococcal infections such as otitis media,
PT   meningitis, and bacteraemia  -
XX
PS   Example 3; Fig 5; 113pp; English.
XX
CC   The invention describes an isolated polypeptide (I) with 70-90%
CC   identity to Streptococcus pneumonia protein BVH-3, BVH-11, variants of
CC   BVH-3 or BVH-11, or chimeric sequences derived from them. A vaccine (II)
CC   comprising (I) is useful for therapeutic or prophylactic treatment of
CC   meningitis, otitis media, bacteraemia or pneumonia infection in an
CC   individual susceptible to these disorders. (II) is also useful for
CC   therapeutic or prophylactic treatment of any streptococcal bacterial
CC   infection (e.g., caused by Streptococcus pneumoniae, group A
CC   Streptococcus such as Streptococcus pyogenes, group B Streptococcus such
CC   as Streptococcus agalactiae, S. dysgalactiae, S. uberis, S. nocardia or
CC   Staphylococcus aureus) in an individual susceptible to the infection.
CC   A polynucleotide (III) encoding (I) is useful in DNA immunisation
CC   techniques. The Streptococcus polypeptides are useful in a diagnostic
CC   test for S. pneumoniae infection. (III) is useful for designing DNA
CC   probes for use in detecting the presence of Streptococcus in a biological
CC   sample suspected of containing the bacteria. The DNA probes may also be
CC   used for detecting circulating S. pneumonia nucleic acid in a sample for
CC   diagnosing streptococcal infections. This sequence encodes the
CC   Streptococcus pneumoniae protein BVH-11-2, used to create the antigenic
CC   peptides described in the method of the invention.
XX
SQ   Sequence 2639 BP; 889 A; 518 C; 567 G; 665 T; 0 other;

  Query Match             41.1%;  Score 980.8;  DB 24;  Length 2639;
  Best Local Similarity   67.8%;  Pred. No. 2.5e-236;
  Matches 1473;  Conservative    0;  Mismatches  638;  Indels   60;  Gaps    5;

Qy       1 TTCTTACGAGTTGGGACTGTATCAAGCTAGAACGGTTAAGGAAAA---TAATCGTGTTTC 57
Db     173 TTCCTATGAACTTGGTCGTCACCAAGCTGGTCAGGTTAAGAAAGAGTCTAATCGAGTTTC 232

Qy      58 CTATATAGATGGAAAACAAGCGACGCAAAAAACGGAGAATTTGACTCCTGATGAGGTTAG 117
Db     233 TTATATAGATGGTGATCAGGCTGGTCAAAAGGCAGAAAATTTGACACCAGATGAAGTCAG 292

Qy     118 CAAGCGTGAAGGAATCAATGCTGAGCAAATCGTCATCAAGATAACAGACCAAGGCTATGT 177
Db     293 TAAGAGAGAGGGGATCAACGCCGAACAAATTGTTATCAAGATTACGGATCAAGGTTATGT 352

Qy     178 CACTTCACATGGCGACCACTATCATTATTACAATGGTAAGGTTCCTTATGACGCTATCAT 237
Db     353 GACCTCTCATGGAGACCATTATCATTACTATAATGGCAAGGTTCCTTATGATGCCATCAT 412

Qy     238 CAGTGAAGAATTACTCATGAAAGATCCAAACTATAAGCTAAAAGATGAGGATATTGTTAA 297
Db     413 CAGTGAAGAACTTCTCATGAAAGATCCGAATTATCAGTTGAAGGATTCAGACATTGTCAA 472

Qy     298 TGAGGTCAAGGGTGGATATGTTATCAAGGTAGATGGAAAATACTATGTTTACCTTAAGGA 357
Db     473 TGAAATCAAGGGTGGCTATGTGATTAAGGTAGACGGAAAATACTATGTTTACCTTAAAGA 532

Qy     358 TGCTGCCCACGCGGATAACGTCCGTACAAAAGAGGAAATCAATCGACAAAAACAAGAGCA 417
Db     533 TGCGGCCCATGCGGACAATATTCGGACAAAAGAAGAGATTAAACGTCAGAAGCAGGAACA 592

Qy     418 TAGTCAACATCGTGAAGGTGGAACTCCAAGAAACGATGGTGCTGTTGCCTTGGCACGTTC 477
Db     593 CAGTCATAATCAT---------AACTCAAGAGCAGATAATGCTGTTGCTGCAGCCAGAGC 643

Qy     478 GCAAGGACGCTATACTACAGATGATGGTTATATCTTTAATGCTTCTGATATCATAGAGGA 537
Db     644 CCAAGGACGTTATACAACGGATGATGGGTATATCTTCAATGCATCTGATATCATTGAGGA 703

Qy     538 TACTGGTGATGCTTATATCGTTCCTCATGGAGATCATTACCATTACATTCCTAAGAATGA 597
Db     704 CACGGGTGATGCTTATATCGTTCCTCACGGCGACCATTACCATTACATTCCTAAGAATGA 763

Qy     598 GTTATCAGCTAGCGAGTTGGCTGCTGCAGAAGCCTTCCTATCTGGTCGAGGAAATCTGTC 657
Db     764 GTTATCAGCTAGCGAGTTAGCTGCTGCAGAAGCCT---------------------------- 798

Qy     658 AAATTCAAGAACCTATCGCCGACAAAATAGCGATAACACTTCAAGAACAAACTGGGTACC 717
Db     799 --------ATTGGAATGGGAAGCAGGGATCTCGTCCTTCTTCAAGTTCTAGTTATAATGC 850

Qy     718 TTCTGTAAGCAATCCAGGAACTACAAATACTAACACAAGCAACAACAGCAACACTAACAG 777
Db     851 AAATCCAGTTCAACCAAGATTGTCAGAGAACCACAATCTGACTGTCACTCCAACTTATCA 910

Qy     778 TCAAGCAAGTCAAAGTAATGACATTGATAGTCTCTTGAAACAGCTCTACAAACTGCCTTT 837
Db     911 TCA---AAATCAAGGGGAAAACATTTCAAGCCTTTTACGTGAATTGTATGCTAAACCCTT 967

Qy     838 GAGTCAACGACATGTAGAATCTGATGGCCTTGTCTTTGATCCAGCACAAATCACAAGTCG 897
Db     968 ATCAGAACGCCATGTAGAATCTGATGGCCTTATTTTCGACCCAGCGCAAATCACAAGTCG 1027

Qy     898 AACAGCTAGAGGTGTTGCAGTGCCACACGGAGATCATTACCACTTCATCCCTTACTCTCA 957
Db    1028 AACCGCCAGAGGTGTAGCTGTCCCTCATGGTAACCATTACCACTTTATCCCTTATGAACA 1087

Qy     958 AATGTCTGAATTGGAAGAACGAATCGCTCGTATTATTCCCCTTCGTTATCGTTCAAACCA 1017
Db    1088 AATGTCTGAATTGGAAAAACGAATTGCTCGTATTATTCCCCTTCGTTATCGTTCAAACCA 1147

Qy    1018 TTGGGTACCAGATTCAAGGCCAGAACAACCAAGTCCACAACCGACTCCGGAACCTAGTCC 1077
Db    1148 TTGGGTACCAGATTCAAGACCAGAACAACCAAGTCCACAATCGACTCCGGAACCTAGTCC 1207

Qy    1078 AGGCCCGCAACCTGCACCAAATCTTAAAATAGACTCAAATTCTTC------------TTT 1125
Db    1208 AAGTCTGCAACCTGCACCAAATCCTCAACCAGCTCCAAGCAATCCAATTGATGAGAAATT 1267

Qy    1126 GGTTAGTCAGCTGGTACGAAAAGTTGGGGAAGGATATGTATTCGAAGAAAAGGGCATCTC 1185
Db    1268 GGTCAAAGAAGCTGTTCGAAAAGTAGGCGATGGTTATGTCTTTGAGGAGAATGGAGTTTC 1327

Qy    1186 TCGTTATGTCTTTGCGAAAGATTTACCATCTGAAACTGTTAAAAATCTTGAAAGCAAGTT 1245
Db    1328 TCGTTATATCCCAGCCAAGGATCTTTCAGCAGAAACAGCAGCAGGCATTGATAGCAAACT 1387

Qy    1246 ATCAAAACAAGAGAGTGTTTCACACACTTTAACTGCTAAAAAAGAAAATGTTGCTCCTCG 1305
Db    1388 GGCCAAGCAGGAAAGTTTATCTCATAAGCTAGGAGCTAAGAAAACTGACCTCCCATCTAG 1447

Qy    1306 TGACCAAGAATTTTATGATAAAGCATATAATCTGTTAACTGAGGCTCATAAAGCCTTGTT 1365
Db    1448 TGATCGAGAATTTTACAATAAGGCTTATGACTTACTAGCAAGAATTCACCAAGATTTACT 1507

Qy    1366 TGNAAATAAGGGTCGTAATTCTGATTTCCAAGCCTTAGACAAATTATTAGAACGCTTGAA 1425
```

```
Db    1508 TGATAATAAAGGTCGACAAGTTGATTTTGAGGTTTTGGATAACCTGTTGGAACGACTCAA 1567

Qy    1426 TGATGAATCGACTAATAAAGAAAAATTGGTAGATGATTTATTGGCATTCCTAGCACCAAT 1485

Db    1568 GGATGTCTCAAGTGATAAAGTCAAGTTAGTGGATGATATTCTTGCCTTCTTAGCTCCGAT 1627

Qy    1486 TACCCATCCAGAGCGACTTGGCAAACCAAATTCTCAAATTGAGTATACTGAAGACGAAGT 1545

Db    1628 TCGTCATCCAGAACGTTTAGGAAAACCAAATGCGCAAATTACCTACACTGATGATGAGAT 1687

Qy    1546 TCGTATTGCTCAATTAGCTGATAAGTATACAACGTCAGATGGTTACATTTTTGATGAACA 1605

Db    1688 TCAAGTAGCCAAGTTGGCAGGCAAGTACACAACAGAAGACGGTTATATCTTTGATCCTCG 1747

Qy    1606 TGATATAATCAGTGATGAAGGAGATGCATATGTAACGCCTCATATGGGCCATAGTCACTG 1665

Db    1748 TGATATAACCAGTGATGAGGGGGATGCCTATGTAACTCCACATATGACCCATAGCCACTG 1807

Qy    1666 GATTGGAAAAGATAGCCTTTCTGATAAGGAAAAAGTTGCAGCTCAAGCCTATACTAAAGA 1725

Db    1808 GATTAAAAAAGATAGTTTGTCTGAAGCTGAGAGAGCGGCAGCCCAGGCTTATGCTAAAGA 1867

Qy    1726 AAAAGGTATCCTACCTCCATCTCCAGACGCAGATGTTAAAGCAAATCCAACTGGAGATAG 1785

Db    1868 GAAAGGTTTGACCCCTCCTTCGACAGACCACCAGGATTCAGGAAATACTGAGGCAAAAGG 1927

Qy    1786 TGCAGCAGCTATTTACAATCGTGTGAAAGGGGAAAAACGAATTCCACTCGTTCGACTTCC 1845

Db    1928 AGCAGAAGCTATCTACAACCGCGTGAAAGCAGCTAAGAAGGTGCCACTTGATCGTATGCC 1987

Qy    1846 ATATATGGTTGAGCATACAGTTGAGGTTAAAAACGGTAATTTGATTATTCCTCATAAGGA 1905

Db    1988 TTACAATCTTCAATATACTGTAGAAGTCAAAAACGGTAGTTTAATCATACCTCATTATGA 2047

Qy    1906 TCATTACCATAATATTAAATTTGCTTGGTTTGATGATCACACATACAAAGCTCCAAATGG 1965

Db    2048 CCATTACCATAACATCAAATTTGAGTGGTTTGACGAAGGCCTTTATGAGGCACCTAAGGG 2107

Qy    1966 CTATACCTTGGAAGATTTGTTTGCGACGATTAAGTACTACGTAGAACACCCTGACGAACG 2025

Db    2108 GTATAGTCTTGAGGATCTTTTGGCGACTGTCAAGTACTATGTCGAACATCCAAACGAACG 2167

Qy    2026 TCCACATTCTAATGATGGATGGGGCAATGCCAGTGAGCATGTGTTAGGCAAGAAAGACCA 2085

Db    2168 TCCGCATTCAGATAATGGTTTTGGTAACGCTAGTGACCATGTTCGTAAAAATAAGGCAGA 2227

Qy    2086 CAGTGAAGATCCAAATAAGAACTTCAAAGCGGATGAAGAGCCAGTAGAGGAAACACCTGC 2145

Db    2228. CCAAGATAGTAAACCTGATGAAGATAAGGAACATGATGAAGTAAGTGAGCCAACTCACCC 2287

Qy    2146 TGAGCCAGAAG 2156

Db    2288 TGAATCTGATG 2298


Search completed: November 17, 2002, 01:50:41
Job time : 373 secs
```